

**PROTEINS OF THE *E. coli*
PRIMOSOME**

by

Nicholas Patrick John Stamford

A thesis submitted for the degree of
Doctor of Philosophy

Research School of Chemistry
Australian National University

October 1991

ACKNOWLEDGEMENTS

I am indebted to my supervisor Dr M.E. Dole for his constant support and guidance throughout my postgraduate training. It has been a privilege to work with him. I would also like to thank the Research School of Chemistry for the provision of such excellent facilities for research and for contributing to a successful postgraduate programme.

I am also indebted to all the members of the laboratory and the Department who have been such pleasant and helpful people to work with. In particular I would like to thank Phillip Doolan and Penny Lilley for their constant support and advice, and also Chris Elvén, Philip Hendry, Brian Gault and Michael Woodhouse for their sound advice and Mary Arnold, Jeff Crowther, Jennifer Doolan and others.

The work described herein is the author's own work, unless otherwise stated, and was carried out within the Research School of Chemistry, Australian National University, from March 1988 - October 1991. None of the material has been submitted in support of an application for any other degree.

University generously and would like to thank all my friends and family members for making life in Canberra such a pleasant experience. Special mention goes to the Doolans and Roger Nixon who have been a constant source of support and encouragement as well as to my friends and Kylie White.

I am also indebted to the Research School of Chemistry and Phillip for their constant support and advice throughout my many years of study. I would also like to thank Jo and John Bush for their wonderful country hospitality and for their love and devotion to their home. Finally, I wish to acknowledge the constant support of my parents, their encouragement and advice throughout my education and for their love and devotion for such a wonderful child.

I especially dedicate this thesis to my wife for her constant support, tolerance and good humour, her patience, her love and her time in helping me to cope with all my trials, and in sharing her experience.

N.P.J. Stamford

ACKNOWLEDGEMENTS

I am indebted to my supervisor Dr N.E. Dixon for his excellent supervision and guidance throughout my postgraduate training. It has been both a pleasure and a privilege to work with him. I would also like to thank the Research School of Chemistry for the provision of such excellent facilities for research and their commitment to a successful postgraduate programme.

I am also indebted to all the members of the laboratory and the department who have been such pleasant and helpful people to work with. In particular I would like to thank Phillipa Doolan and Penny Lilley for their excellent technical assistance, Jenny Beck, Chris Elvin, Philip Hendry, Brian Surin and Subhash Vasudevan for their sound advice and Mary Argall, Jeff Crowther, Jeremy Garwood and Phillip Thompson for their constant friendship. I also wish to acknowledge the support and advice of my advisers Dr W.L.F Armarego and Dr G.D. Smith.

I have enjoyed my undergraduate and postgraduate years at the Australian National University immensely and would like to thank all my friends and acquaintances for making life in Canberra such a pleasurable experience. Special mentions go to Tony Doulgeris and Roger Nixon who have been a constant source of support and good humour as well as Wesley Phoa and Kylie White.

I am as ever indebted to my brothers Christopher and Roger for their encouragement, support and tolerance throughout my many years of study. I would also like to thank Jo and John Bush for their wonderful country hospitality during my all too brief visits to their home. Finally, I wish to acknowledge the continued support of my parents, their encouragement and advice throughout my education and for their love and tireless devotion for such a hopeless child.

I especially dedicate this thesis to my wife for her constant support, tolerance and good humour, her patience, her love and her time in helping me to cope with all my trials, and in sharing this experience.

PUBLICATIONS

Evins, C.M., Thompson, P.R., Argall, M.E., Hendry, P., Stanford, N.P.J., Lilley, P.E. and Dixon, N.E. (1991) Modified bacteriophage lambda promoters for the overproduction of proteins in *Escherichia coli*. *Gene* 107: 123-126.

Stanford, N.P.J., Lilley, P.E. and Dixon, N.E. (1991) High-level overproduction of *Escherichia coli* structural proteins. *Proc. Aust. Biochem. Soc.* 21, 57-61.

Dixon, N.E., Argall, M.E., Rock, J.L., Evins, C.M., Garwood, J.J., Hendry, P., Lilley, P.E., Stanford, N.P.J. and Thompson, P.R. (1989) A versatile new series of lambda phage vectors for overproduction of proteins in *Escherichia coli*. *Proc. Aust. Biochem. Soc.* 21, P60.

Stanford, N.P.J. and Dixon, N.E. (1991) Overproduction, purification and properties of the *Escherichia coli* dnaB and dnaC proteins and the dnaB-dnaC complex. *Proc. Aust. Biochem. Soc.* 22, 10-14.

To

Linda Anne

Stanford, N.P.J., Lilley, P.E. and Dixon, N.E. (1991) Overproduction, purification and properties of the *Escherichia coli* primase. *Proc. Aust. Biochem. Soc.* 22, P1.

Stanford, N.P.J., Lilley, P.E. and Dixon, N.E. Purified copies of *Escherichia coli* replication proteins. *Excerpta* primase is a zinc metallo-protein. Submitted for publication (1991).

Stanford, N.P.J. and Dixon, N.E. Strategies for improved overproduction of the *Escherichia coli* dnaB and dnaC proteins. In preparation.

Stanford, N.P.J., Lewis, P.J., Wake, R.G. and Dixon, N.E. Purification and properties of the *Escherichia coli* dnaB and dnaC proteins, and the dnaB-dnaC complex. In preparation.

Lilley, P.E., Stanford, N.P.J., Vasudevan, S.G. and Dixon, N.E. Physical map of the *Escherichia coli* chromosome in the region between the *ori* operon and the *oriC* to *oriD* genes. Chromosomal location of *oriC*. In preparation.

PUBLICATIONS

Elvin, C.M., Thompson, P.R., Argall, M.E., Hendry, P., Stamford, N.P.J., Lilley, P.E. and Dixon, N.E. (1990) Modified bacteriophage lambda promoter vectors for overproduction of proteins in *Escherichia coli*. *Gene* **87**, 123-126

Stamford, N.P.J., Lilley, P.E. and Dixon, N.E. (1989) High-level overproduction of *Escherichia coli* primosomal proteins. *Proc. Aust. Biochem. Soc.* **21**, SP21

Dixon, N.E., Argall, M.E., Beck, J.L., Elvin, C.M., Garwood, J.J., Hendry, P., Lilley, P.E., Stamford, N.P.J. and Thompson, P.R. (1989) A versatile new series of λ promoter vectors for overproduction of proteins in *Escherichia coli*. *Proc. Aust. Biochem. Soc.* **21**, P60

Stamford, N.P.J. and Dixon, N.E. (1991) Overproduction, purification and properties of the *Escherichia coli* dnaB and dnaC proteins and the dnaB/dnaC complex. *Proc. Aust. Biochem. Soc.* **23**, SP10

Stamford, N.P.J., Lilley, P.E. and Dixon, N.E. (1991) Overproduction, purification and properties of the *Escherichia coli* primase. *Proc. Aust. Biochem. Soc.* **23**, P11

Stamford, N.P.J., Lilley, P.E. and Dixon, N.E. Enriched sources of *Escherichia coli* replication proteins. The dnaG primase is a zinc metalloprotein.

Submitted for publication (1991)

Stamford, N.P.J. and Dixon, N.E. Strategies for improved overproduction of the *Escherichia coli* dnaB and dnaC proteins. In preparation.

Stamford, N.P.J., Lewis, P.J., Wake, R.G. and Dixon, N.E. Purification and properties of the *Escherichia coli* dnaB and dnaC proteins, and the dnaB-dnaC complex. In preparation.

Lilley, P.E., Stamford, N.P.J., Vasudevan, S.G. and Dixon, N.E. Physical map of the *Escherichia coli* chromosome in the region between the *mal* operon and *ssb* (91.5 to 92 min). Chromosomal location of *alr*. In preparation.

ABBREVIATIONS

A_{λ}	- absorbance at wavelength λ
bis-acrylamide	- N,N'-methylene-bis-acrylamide
bp	- base pair(s)
Brij-58	- polyoxyethylene 20 cetyl ether
BSA	- bovine serum albumin
DNA pol I	- DNA polymerase I
dNTP(s)	- deoxyribonucleoside triphosphate(s)
DTT	- dithiothreitol
EDTA	- ethylenediamine tetraacetic acid disodium salt
FPLC	- fast protein liquid chromatography
holoenzyme	- DNA polymerase III holoenzyme
IHF	- integration host factor
kb	- kilobase pairs
kDa	- kilodaltons
LB	- Luria-Bertani medium
MES	- 2-(N-morpholino)ethanesulfonic acid
M_r	- relative molecular weight
NaPP _i	- Sodium pyrophosphate
NEM	- N-ethylmaleimide
ntd(s)	- nucleotide(s)
ORF	- open reading frame
PAGE	- polyacrylamide gel electrophoresis
pfu	- plaque forming units
P_L	- major leftward promoter of bacteriophage λ
P_R	- major rightward promoter of bacteriophage λ
RBS	- ribosome-binding site
RF	- replicative form
RNAP	- RNA polymerase holoenzyme
RNaseH	- ribonuclease H
rNTP	- ribonucleoside triphosphate(s)
SDS	- sodium dodecyl sulfate
SS	- single-strand(ed)
TEMED	- N,N,N',N'-tetramethylethylenediamine
Tris	- tris(hydroxymethyl) aminomethane
ts	- temperature sensitive
UV	- ultraviolet

ABSTRACT

The *Escherichia coli* primosome is a complex of seven proteins - primase, the dnaB and dnaC proteins, and the products of the *dnaT*, *priA*, *priB* and *priC* genes. It functions as a mobile replication promoter by laying down short RNA primers for DNA synthesis during complementary strand replication of bacteriophage ϕ X174 DNA, in replication of plasmid pBR322, and on the lagging strand at replication forks in chromosomal DNA replication. Work presented in this thesis encompasses several aspects of the protein-protein and protein-nucleic acid interactions that occur with these proteins during the initiation of replication of the *E. coli* chromosome and the synthesis of RNA primers required for DNA replication.

The primosomal proteins are present in wild-type *E. coli* strains in relatively low cellular abundance. To undertake studies into their structure and function it was first necessary to provide enriched sources of the proteins to facilitate their isolation in large quantities. This was achieved for primase and the dnaB, dnaC and dnaT proteins by replacing the DNA sequences responsible for the normal control of their expression. In each case the natural ribosome-binding site of the gene was removed by treatment of an appropriate DNA fragment with the double-stranded exonuclease Bal31. Alternative strategies were then employed to provide a new ribosome-binding site perfectly complementary to the 3'-OH terminus of *E. coli* 16-S ribosomal RNA, variably spaced upstream of the ATG translation initiation codon, and to place transcription of the gene under the control of tandem strong bacteriophage λ promoters P_R and P_L in the high copy number vector pCE30, or a derivative of it (Elvin *et al.*, 1990). These vectors also direct expression of the λ cI857^{ts} gene product, allowing control of transcription of a gene by shift in the temperature of the culture from 30 °C to 42 °C.

The overproduced *dnaG* primase was soluble and could be extracted almost quantitatively into a cleared cell lysate. Primase from this source was used for the purification of quantities of this enzyme sufficient for chemical and structural studies and was shown to contain a stoichiometric amount of zinc.

The proteins produced following overexpression of the *dnaB*, *dnaC* and *dnaT* genes were completely insoluble. Novel strategies were developed to solubilize the dnaB and dnaC protein. By overproducing both of these proteins in a single cell, it was possible to render them soluble, indicating a direct interaction between them *in vivo*.

Both the *dnaC* and *dnaB* proteins could be extracted almost quantitatively into a cleared cell lysate. The application of DEAE-cellulose anion-exchange chromatography enabled the isolation and complete separation of not only the *dnaB* and *dnaC* proteins, but also an intact *dnaB*-*dnaC* complex, each in quantities sufficient for chemical and structural studies. The ratio of specific activities of the purified proteins suggested a B6:C6 complex, and this stoichiometry of *dnaB* and *dnaC* in the complex was confirmed by quantitative electrophoretic analysis and gel filtration.

Analysis of DNA sequences downstream of the *dnaB* gene revealed a second open reading frame. This potential coding region was not terminated within that DNA fragment and was largely homologous to the 5' terminus of the *alr* gene from *S. typhimurium* which encodes D-alanine racemase. It is likely that in both *E. coli* and *S. typhimurium* the *dnaB* and *alr* genes together form an operon.

Examination of the sequences of several *dnaB* homologues revealed a highly-conserved leucine heptad repeat sequence near the COOH-terminus. The possibility that this motif is involved in protomer association in the *dnaB* hexamer was investigated in a preliminary way by examining the effects of (conservative) substitution of valine residues for heptad repeat leucines. The proteins were produced in a *dnaB*⁺ strain using mutant derivatives of a plasmid that directed high-level overproduction of soluble wild-type *dnaB* protein, purified, and examined for size and activity. Although most of the mutants behaved like the wild-type, two proteins with multiple substitutions were largely insoluble and had reduced activities. Estimation of the molecular weights of the purified mutant *dnaB* proteins determined that one of these appeared to be a mixture of oligomeric species.

The *dnaB* protein was also analysed by overproducing genetically engineered polypeptides that were thought to approximate discrete structural domains within the wild-type protein. This was achieved using restriction endonucleases to generate fragments of the *dnaB* gene which were then expressed in *E. coli*. Although polypeptides thought to mimic the NH₂-terminal domain of *dnaB* were soluble, the COOH-terminal fragments were not. This suggests that the COOH-terminus is responsible for the insolubility of overproduced wild-type *dnaB* protein and that the *dnaC* protein interacts with this domain.

The gene of a *Chlamydia trachomatis* homologue of the *E. coli* *dnaB* protein was also manipulated. Although plasmids were isolated that were predicted to direct the overproduction of the *dnaB*-like protein, none was detected. Possible reasons for the failure of *E. coli* to express the *C. trachomatis* gene are discussed.

TABLE OF CONTENTS

	<u>Page</u>
DECLARATION	i
ACKNOWLEDGEMENTS	ii
DEDICATION	iii
PUBLICATIONS	iv
ABBREVIATIONS	v
ABSTRACT	vi
 CHAPTER 1: GENERAL INTRODUCTION	 1
 CHAPTER 2: GENERAL MATERIALS AND METHODS	 13
2.1 Bacterial Strains and Plasmid Vectors	14
2.2 Chemicals, Reagents and Instruments	14
2.3 Protocols for Molecular Cloning	15
2.4 Protocols for Protein Manipulation	24
 CHAPTER 3: IMPROVED OVERPRODUCTION AND PURIFICATION OF THE <i>Escherichia coli</i> <i>dnaG</i> PRIMASE, A ZINC METALLOPROTEIN	 27
3.1 Introduction	27
3.2 Materials and Methods	29
3.3 Results	32
3.4 Discussion	35
 CHAPTER 4: IMPROVED OVERPRODUCTION OF THE <i>dnaB</i> , <i>dnaC</i> AND <i>dnaT</i> PROTEINS	 41
4.1 Introduction	41
4.2 Materials and Methods	42
4.3 Results	47
4.4 Discussion	56

**CHAPTER 5: PURIFICATION AND PROPERTIES OF THE dnaB
AND dnaC PROTEINS AND A dnaB-dnaC COMPLEX**

5.1	Introduction	62
5.2	Materials and Methods	63
5.3	Results	66
5.4	Discussion	72

**CHAPTER 6: INVOLVEMENT OF A LEUCINE HEPTAD REPEAT
MOTIF IN OLIGOMERIZATION OF dnaB PROTEIN**

6.1	Introduction	76
6.2	Materials and Methods	79
6.3	Results	83
6.4	Discussion	86

CHAPTER 7: DOMAIN STRUCTURE OF THE dnaB PROTEIN

7.1	Introduction	93
7.2	Materials and Methods	94
7.3	Results	96
7.4	Discussion	99

CONCLUDING STATEMENT	102
APPENDIX A	104
APPENDIX B	110
REFERENCES	115

CHAPTER 1

General Introduction

The process of chromosomal replication has an essential role in the cell cycle of all organisms. If the full complement of genetic material is not faithfully duplicated then an accurate copy cannot be received by each of the new daughter cells. Furthermore, the crucial processes of genetic rearrangement and repair of the genome ultimately rely on the replication machinery of the cell. The bacterium *Escherichia coli*, a gram-negative rod of the family *Enterobacteriaceae*, has a double-stranded circular chromosome which consists of approximately 4.7×10^6 base pairs (bp). The replication of this genome starts at a unique origin (*oriC*) and proceeds bidirectionally by a semiconservative mechanism whereby each strand serves as a template for a new complementary strand. The work presented in this thesis encompasses several aspects of protein-protein and protein-nucleic acid interactions that occur during the initiation of replication of the *E. coli* chromosome and the synthesis of RNA primer transcripts that mark the commencement of DNA synthesis.

The current understanding of DNA replication is based on the identification and genetic analysis of replication components, and determination of their function primarily through *in vitro* replication of various small *E. coli* bacteriophages, which rely predominantly on host encoded proteins, and small multicopy plasmids that carry *oriC* as the sole replication origin. From these systems the overall mechanism of DNA replication was established and can be viewed as a process comprised of three discrete phases: (i) Initiation: the tightly regulated process by which elongation complexes are created at a unique origin on double-stranded DNA; (ii) Elongation: the coordinated synthesis of the complementary strand by a protein complex at the replication fork progressing along the chromosome; and (iii) Termination: the resolution and segregation of the two newly replicated daughter chromosomes (for review see Kornberg and Baker, 1991).

Initiation

The replication of the *E. coli* chromosome is strictly regulated at the stage of initiation (von Meyenburg and Hansen, 1987). While considerable information is available regarding the physical initiation process at *oriC*, the mechanism regulating this and the tightly coupled process of cell division remain unclear. The signal that triggers the replication process is thought to be related to the ratio of cell mass with a specific protein (D'Arai and Boulloc, 1990; Zyskind, 1990). One such protein could be *dnaA*, a protein that binds sequences in *oriC* in the first step of chromosomal replication and whose concentration appears to be a critical factor in determining the timing of initiation in the cell cycle (Atlung *et al.*, 1987; Løbner-Olesen *et al.*, 1989).

Cellular levels of *dnaA* protein are autoregulated. Transcription of *dnaA* initiates at either of two promoters, both of which are repressed by the *dnaA* protein itself (Braun *et al.*, 1985). However, the levels of *dnaA* present in *E. coli* are probably too large for the binding of *dnaA* protein to *oriC* to be the sole mechanism of initiation control (Sekimizu *et al.*, 1988). On the other hand close association of cell membrane with an *oriC* complex (Firshein, 1989; Norris, 1990) and the *dnaA* protein (Hwang *et al.*, 1990) may be essential for the successful regulation and initiation of chromosomal replication.

The cell membrane may directly regulate the level of active monomeric *dnaA* within the cell by localizing *dnaA* protein in an inert aggregated phospholipid-bound form (Hwang *et al.*, 1990). The *dnaA* protein-membrane association may also function to coordinate initiations within a single cell through the activation of the protein. Initiation of replication at *oriC* is specifically dependent on an ATP-bound form of *dnaA*. The native *dnaA* protein and the ADP-bound form of *dnaA*, while still able to bind sequences within the origin, are effectively inactive in initiating replication (Sekimizu *et al.*, 1987). The importance of *dnaA* association with the membrane is therefore thought to be related to the ability of acidic phospholipids (in particular cardiolipin) to promote the exchange of bound ADP for ATP, ensuring that *dnaA* remains active during the initiation process (Sekimizu and Kornberg, 1988; Yung and Kornberg, 1988). Other mechanisms which have been proposed for control of replication include activation of the *dnaA* protein by cAMP (Hughes *et al.*, 1988), interaction of *dnaA* with other proteins (Hwang and Kaguni, 1991), inhibition of replication initiation through a protein bound at the origin (Hwang and Kornberg, 1990), methylation of GATC sites within *oriC* by *dam* methyltransferase (Boye and Løbner-Olesen, 1990; Campbell and Kleckner, 1990), and transcription of a gene close to *oriC* (*mioC*) which is in turn repressed strongly by the *dnaA* protein (Lothar *et al.*, 1985; Stuitje *et al.*, 1986).

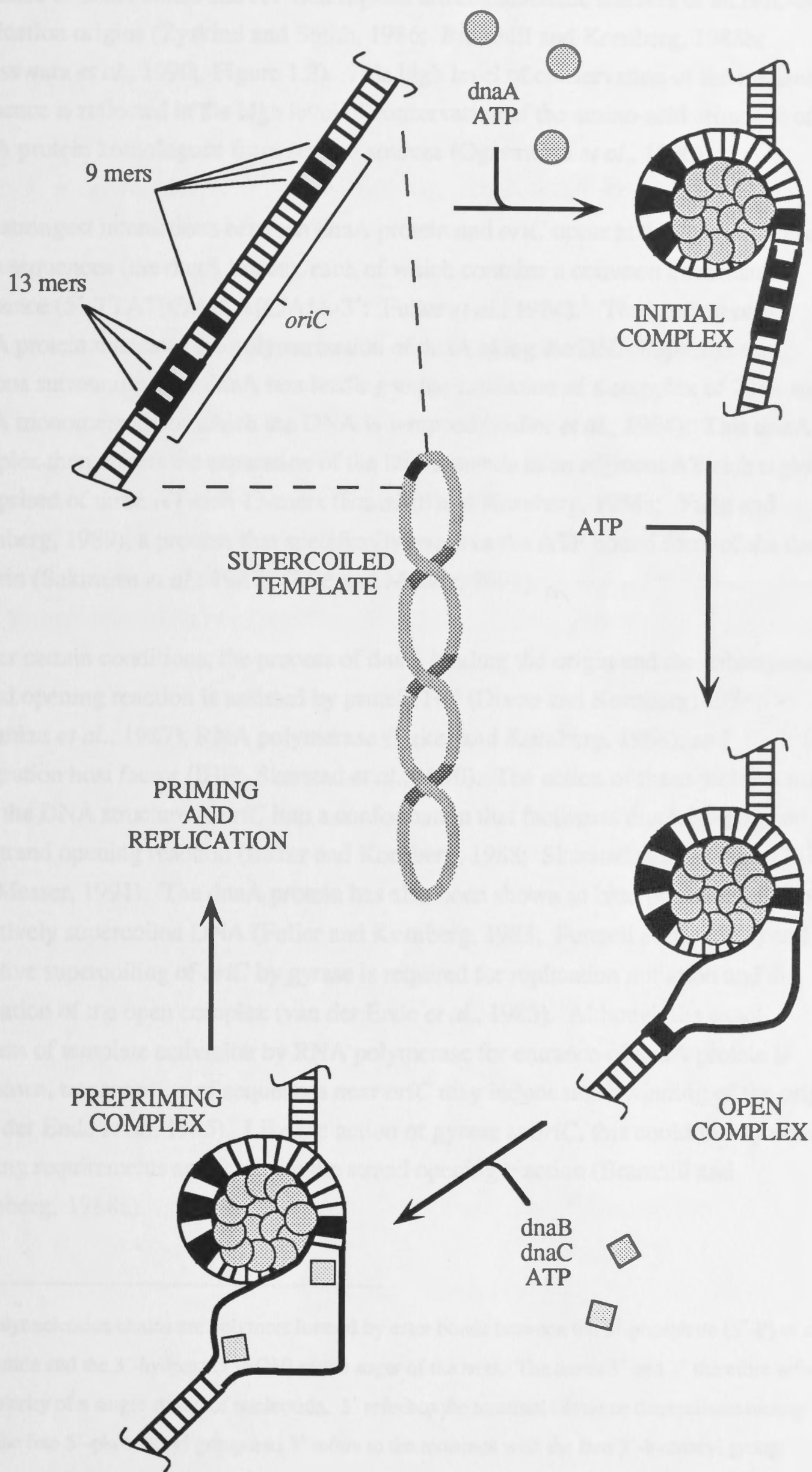
A proposed model for initiation at *oriC* identifies the *dnaA* protein as performing three crucial roles: (i) it binds tightly to four 9-bp repeat sequences in the origin (termed "dnaA boxes") to form an initial *dnaA-oriC* complex; (ii) it 'melts' adjacent AT-rich segments of DNA to form an open complex; and (iii) it guides the *dnaB-dnaC* complex into this opened region to form a prepriming complex which marks the future forks of bidirectional DNA replication (Bramhill and Kornberg, 1988a; Figure 1.1).

The *E. coli* origin of replication (*oriC*) is a 245-bp segment of the chromosome (Oka *et al.*, 1980) and the *dnaA* protein binds specifically to the *dnaA* boxes within this origin. *OriC* shares considerable similarity with the replication origins of other gram-negative

Figure 1.1

A scheme for initiation at *oriC*. The *dnaA* protein binds to the four 9-mers to organize *oriC* around a protein core to form the initial complex. The three 13-mers are then melted serially by *dnaA* protein to create the open complex. The *dnaB*-*dnaC* complex can now be directed to the 13-mer region, resulting ultimately in more extensive duplex opening and generation of a prepriming complex.

Figure adapted from Bramhill and Kornberg (1988a).



and gram-positive organisms at the level of genetic arrangement and sequences. The presence of *dnaA* boxes and AT-rich regions are characteristic features of all *oriC*-like replication origins (Zyskind and Smith, 1986; Bramhill and Kornberg, 1988b; Ogasawara *et al.*, 1990; Figure 1.2). This high level of conservation of the nucleotide sequence is reflected in the high level of conservation of the amino-acid sequence of *dnaA* protein homologues from several sources (Ogasawara *et al.*, 1990).

The strongest interactions between *dnaA* protein and *oriC* occur at the four conserved 9-bp sequences (the *dnaA* boxes), each of which contains a common consensus sequence (5'-TTAT[C/A]CA[C/A]A-3'; Fuller *et al.*, 1984).¹ The binding of the *dnaA* protein nucleates the polymerization of *dnaA* along the DNA duplex in the regions surrounding the *dnaA* box leading to the formation of a complex of 20 or more *dnaA* monomers about which the DNA is wrapped (Fuller *et al.*, 1984). This *dnaA* complex then effects the separation of the DNA strands in an adjacent AT-rich region comprised of three AT-rich 13-mers (Bramhill and Kornberg, 1988a; Yung and Kornberg, 1989), a process that specifically requires the ATP bound form of the *dnaA* protein (Sekimizu *et al.*, 1987; Gille and Messer, 1991).

Under certain conditions, the process of *dnaA* binding the origin and the subsequent strand opening reaction is assisted by protein HU (Dixon and Kornberg, 1984; Sekimizu *et al.*, 1987), RNA polymerase (Baker and Kornberg, 1988), and integration host factor (IHF; Skarstad *et al.*, 1990). The action of these proteins may alter the DNA structure at *oriC* into a conformation that facilitates *dnaA* binding and the strand opening reaction (Baker and Kornberg, 1988; Skarstad *et al.*, 1990; Gille and Messer, 1991). The *dnaA* protein has also been shown to bind preferentially to negatively supercoiled DNA (Fuller and Kornberg, 1983; Funnell *et al.*, 1987) and negative supercoiling of *oriC* by gyrase is required for replication initiation and the formation of the open complex (van der Ende *et al.*, 1985). Although the exact process of template activation by RNA polymerase for entrance of *dnaA* protein is unknown, transcription of sequences near *oriC* may induce underwinding of the origin (van der Ende *et al.*, 1985). Like the action of gyrase at *oriC*, this could satisfy *dnaA* binding requirements and facilitate the strand opening reaction (Bramhill and Kornberg, 1988a).

1. Polynucleotide chains are polymers formed by ester bonds between the 5'-phosphate (5'-P) of one nucleotide and the 3'-hydroxyl (3'-OH) of the sugar of the next. The terms 5' and 3' therefore refer to the polarity of a single strand of nucleotide. 5' refers to the terminal ribose or deoxyribose moiety with the free 5'-phosphoryl group and 3' refers to the terminus with the free 3'-hydroxyl group.

Figure 1.2

Consensus sequence of the minimal origin of enteric bacterial chromosomes. The numbering of the nucleotide positions is that used for *E. coli*, and the 5' end is at the upper left. Capital letter = the same nucleotide in all five origins; lower case letter = the same nucleotide in three of the five bacterial origins, with only two different nucleotides at the site; n = any of the four possible nucleotides, or a deletion (-); underline = GATC *dam* methylation sites. The 245-bp minimal origin of *E. coli* is enclosed in brackets; the three AT-rich 13-mers, L (left), R (right), and M (middle), are indicated by lightly shaded boxes; and the four related 9-bp dnaA binding sequences, R1, R2, R3, and R4 (dnaA boxes) are indicated by underlined heavily shaded boxes. A related sequence, which may represent a fifth dnaA box, R5, is found between *E. coli* nucleotide positions 135 and 143. DNA unwinding is initiated at the right-hand 13-mer (Bramhill and Kornberg, 1988a) and bidirectional synthesis is initiated between dnaA boxes R2 and R4 in the right part of *oriC* (Seufert and Messer, 1987).

Figure adapted from Kornberg and Baker (1991).

The absolute dependence on single-stranded binding protein (SSB) during the initiation stage of replication from *oriC* indicates the formation of an open DNA structure (van der Ende *et al.*, 1985). The second stage of initiation of chromosomal replication requires the delivery of the *dnaB* protein to this open complex to extend strand separation at the origin enabling the entry and assembly of components of the DNA replication complex ("replisome"). The *dnaB* protein is the major replicative helicase in *E. coli* whose function is to unwind the DNA duplex ahead of the replication complex. The *dnaB* protein alone is unable to bind single-stranded DNA complexed with SSB (LeBowitz and McMacken, 1986; Baker *et al.*, 1986) and the action of *dnaA* protein at the replication origin is therefore ultimately required for the delivery of the *dnaB* protein to this SSB-covered template (Funnell *et al.*, 1987; Bramhill and Kornberg, 1988a).

Delivery of the *dnaB* protein to the *dnaA-oriC* complex requires its prior interaction with *dnaC* protein (Wahle *et al.*, 1989b), a protein with which *dnaB* forms a tight isolable complex in the presence of ATP (Kobori and Kornberg, 1982c; Wahle *et al.*, 1989a; Figure 1.2). The absolute dependence of the *dnaA*-dependent initiation reaction on *dnaC* may in part be explained by protein-protein interactions that enable *dnaB* to gain access to the SSB-coated template. Either the *dnaB* protein forms a unique conformation in the *dnaB-dnaC* complex that enables it to recognize and bind particular structures at the *E. coli* origin or alternatively, *dnaC* may interact directly with proteins in the origin complex (Wahle *et al.*, 1989b). The interaction of the *dnaB-dnaC* complex with the origin complex leads to the ATP-dependent release of the *dnaC* protein (Wahle *et al.*, 1989b), a process required for the normal function of *dnaB*.

An analogous reaction to the *dnaC*-dependent delivery of the *dnaB* protein to *oriC* has also been observed in the replication of phage λ (McMacken *et al.*, 1986; Dodson *et al.*, 1989; Zylicz *et al.*, 1988). In this system, the phage-encoded initiation protein O binds *ori λ* sequences and induces the formation of single-stranded open complexes near *ori λ* (Schnoss *et al.*, 1988; Alfano and McMacken, 1989). The phage P protein, like *dnaC*, forms a complex with the host *dnaB* protein (McMacken *et al.*, 1983; Zylicz *et al.*, 1984) and through interactions with λ O protein bound to the λ origin, delivers the *dnaB* protein to the template (Dodson *et al.*, 1985). The λ P protein is a potent inhibitor of *dnaB* function (Dodson *et al.*, 1986; Liberek *et al.*, 1988), and several host-encoded proteins (*dnaK*, *dnaJ* and *grpE*) are required to liberate λ P protein from the origin complex to enable *dnaB* to function (Zylicz *et al.* 1989; Mallory *et al.*, 1990).

The λ O and λ P proteins can also assemble the dnaB protein onto SSB-coated single-stranded DNA (LeBowitz *et al.*, 1985). In this case *ori* λ is not required and it is the strong affinity of the λ O protein for single-stranded DNA that probably initiates template interactions. A similar system has also been observed for the dnaA and dnaC-dependent delivery of dnaB onto single-stranded DNA (Wahle *et al.*, 1989b; Masai *et al.*, 1990b). This process also operates in the absence of *ori*C and suggests that the complex formed during replication initiation at *ori*C and *ori* λ is not specifically required for delivery of dnaB to the template but rather for the strand opening reactions.

Once the dnaB protein is properly positioned at *ori*C as a consequence of its interactions with the dnaA and dnaC proteins, it proceeds to unwind the duplex DNA in the vicinity of the *E. coli* origin (Baker *et al.*, 1986). Single-stranded binding protein then binds to the separated complementary strands created by the helicase activity of dnaB protein, thereby stopping their rapid reassociation. Melting of the DNA duplex proceeds bidirectionally from *ori*C as the dnaB helicase migrates from the dnaA-*ori*C complex to provide a template for the priming and replication enzymes (Baker *et al.*, 1987). The fate of the dnaA protein bound at the origin is not known, although it may well be displaced by the helicase action of dnaB.

Elongation

The second crucial process in DNA replication is the rapid and accurate reproduction of the parent chromosome. The enzyme that performs this polynucleotide synthesis is the dimeric DNA polymerase III holoenzyme (holoenzyme), the major replicative DNA polymerase of *E. coli* (McHenry, 1988). The holoenzyme is a large, highly-processive asymmetric multisubunit complex (~900 kDa) comprised of 20 or more polypeptides (Maki *et al.*, 1988). It is responsible not only for high-fidelity synthesis of most of the *E. coli* chromosome but also for the proof-reading of the newly generated strand (McHenry, 1988; Maki *et al.*, 1988). The holoenzyme can catalyse DNA synthesis at a rate of ~500 ntds/sec at 30 °C during replication on singly primed ϕ X174 single-stranded DNA *in vitro* (O'Donnell and Kornberg, 1985) and possesses a processivity in excess of 5000 nucleotides (Fay *et al.*, 1981, 1982). Two basic features of all DNA polymerases, however, define the process by which holoenzyme actions are initiated and ultimately how the chromosome is replicated. Firstly, unlike RNA polymerases, DNA polymerases cannot initiate polynucleotide synthesis but can only extend chains that have the 3'-OH terminus of a nucleotide paired to a template strand extended beyond it. Secondly, as a consequence of their chemistry that

involves promotion of nucleophilic attack by a 3'-OH group on the α -phosphate of a dNTP, DNA polymerase can extend chains only in the 5' \rightarrow 3' direction.

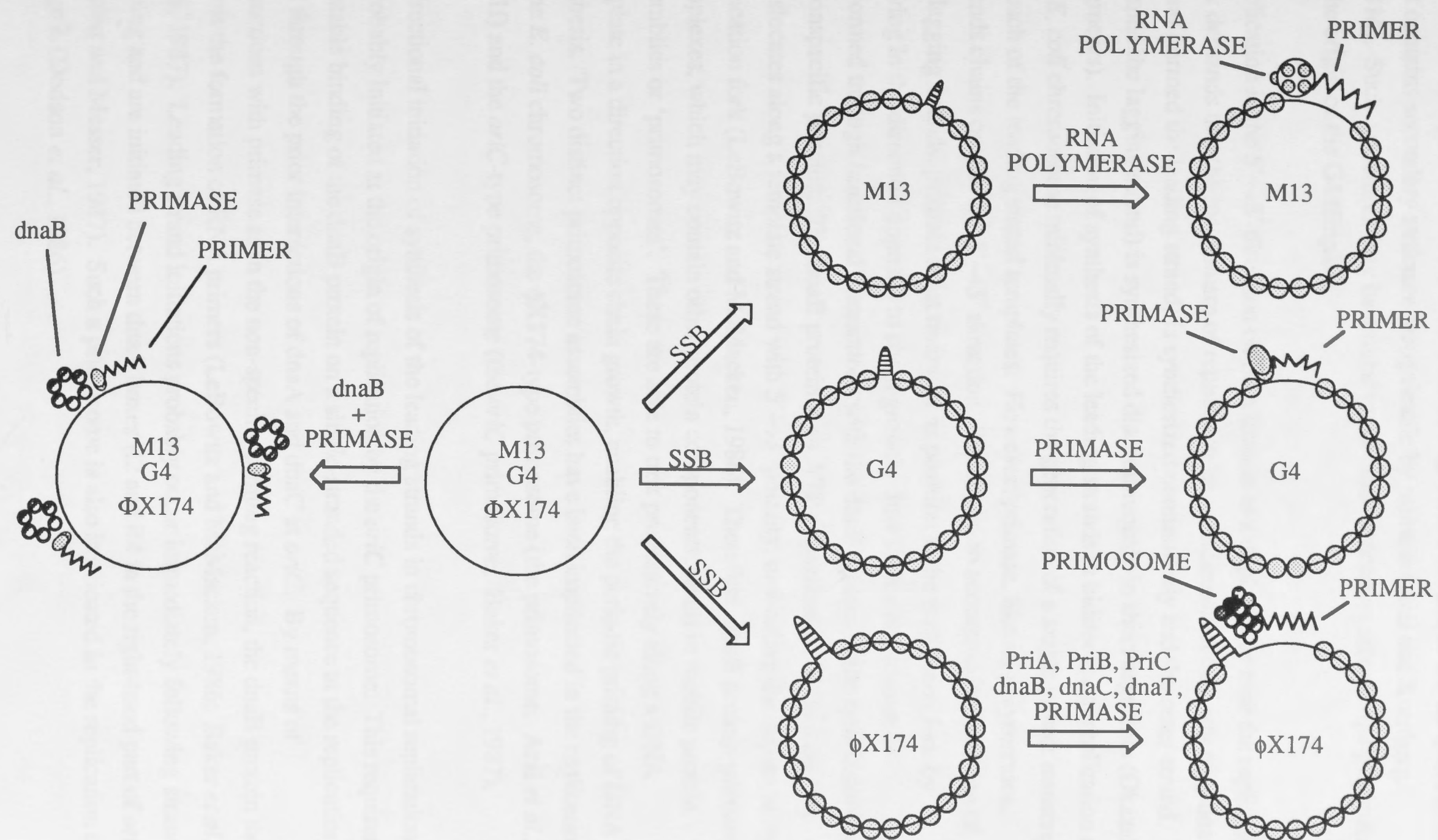
Except for several viral systems that involve priming by terminal proteins (Salas, 1988; Challberg and Kelly, 1989) a short RNA transcript (RNA primer) is the most general device used in initiating the covalent start by a DNA polymerase. The use of RNA/DNA hybrids for initiation may be to ensure a high fidelity in DNA replication, as the proof-reading and error correcting mechanisms in DNA replication may not function as well in removing base-pairing errors at or near the start of a DNA chain as during chain growth. In *E. coli* the initiation of DNA synthesis through the formation of an RNA primer is strictly dependent on the specialized RNA polymerase, primase (van der Ende *et al.*, 1985). Studies of primer RNA synthesis on prokaryotic single-stranded phage templates has revealed four separate modes of priming DNA synthesis (Figure 1.3). Although RNA polymerase holoenzyme can also serve to generate primers only a few RNA priming events depend on this enzyme. These include the generation of primers on templates of the filamentous phages and some plasmids (*e.g.* ColE1). Transcription by RNA polymerase has been implicated as an early and key event in replication of the *E. coli* chromosome (Zyskind and Smith, 1977), and substantial DNA synthesis occurring in the absence of added primase in the *in vitro* replication of *oriC* plasmids demonstrates that RNA polymerase can directly prime replication at *oriC* (Kaguni and Kornberg, 1984). The process is far less efficient than the primase catalysed reaction (Ogawa *et al.*, 1985) and, *in vivo*, transcription by RNA polymerase probably contributes by activating the initiation process at *oriC* (termed "transcriptional activation", see Baker and Kornberg, 1988).

Elucidation of the properties of the *E. coli* primase has largely depended on its function at the complementary-strand origin of phage G4. Primase can bind and transcribe the G4 template unaided but first requires the formation of specific SSB-dependent secondary structures on the single-stranded template (Lambert *et al.*, 1986; Hiasa *et al.*, 1990). Primase function on most other templates, however, requires the addition of the multifunctional dnaB protein. A distributive priming mechanism that operates on all single-stranded templates upon which primase is normally inactive in the absence of single-stranded binding protein has been termed general 'nonspecific' priming (Arai and Kornberg, 1979). In this scheme priming depends on only dnaB and dnaG primase, suggesting a functional interaction between the two proteins. The dnaB protein and the primase alone associate with the single-stranded template to manufacture an RNA primer for initiation of DNA synthesis (Arai and Kornberg, 1981d). The precise mechanism by which the dnaB protein induces primase to synthesize primer transcripts has not been defined and there is so far no genetic or

Figure 1.3

Primer RNA synthesis on prokaryotic single-stranded phages has revealed four different modes of priming: (i) Host RNA polymerase-synthesized primers at a defined region on SSB-coated M13 single-stranded DNA (Geider *et al.*, 1978), (ii) Primase-synthesized RNA at the complementary strand origin of SSB-coated G4 phage DNA (Bouché *et al.*, 1978); (iii) Primer RNA synthesis on SSB-coated ϕ X174 DNA, requiring the action of seven prepriming proteins (Shlomain *et al.*, 1981); and (iv) Primase-catalysed primer RNA synthesis on uncoated single-stranded DNA in the presence of dnaB protein (general or nonspecific priming; Arai and Kornberg, 1981d).

Figure adapted from Kornberg and Baker (1991).



biochemical evidence to suggest a protein-protein interaction between primase and *dnaB*. One proposal is that the *dnaB* protein engineers a feature in the template strand that contains secondary structure recognizable by primase (Arai and Kornberg, 1981d). Such structures may be similar to the SSB-dependent stem-loops generated at the origin of the G4 template.

Replication in the 5'→3' direction on both strands of a duplex at or near the replication fork demands that the mechanism of replication be semidiscontinuous, whereby one strand (termed the leading strand) is synthesized continuously and the other strand (termed the lagging strand) is synthesized discontinuously in short segments (Okazaki fragments). Initiation of synthesis of the leading strands in bidirectional replication of the *E. coli* chromosome minimally requires the generation of a single primer transcript on each of the leading strand templates. However, primase, like all polymerases, extends chains only in the 5'→3' direction. Therefore, to accommodate priming of the lagging strands, primase must maintain its position at the replication fork by moving in the direction opposite to chain growth. Just such a mechanism is performed through functional interactions with the *dnaB* protein of the type identified in nonspecific priming. The *dnaB* protein is an ATP-dependent DNA helicase and translocates along a template strand with 5'→3' polarity, unwinding the duplex at the replication fork (LeBowitz and McMacken, 1986). Therefore, *dnaB* protein-primase complexes, which may contain other protein components, result in mobile protein assemblies or 'primosomes'. These are able to trek processively along a DNA template in a direction opposite chain growth, enabling the periodic priming of DNA synthesis. Two distinct primosome assemblies have been implicated in the replication of the *E. coli* chromosome, the ϕ X174-type primosome (the primosome; Arai *et al.*, 1981f) and the *oriC*-type primosome (the *oriC* primosome; Baker *et al.*, 1987).

Bidirectional initiation of synthesis of the leading strands in chromosomal replication is probably initiated at the origin of replication by the *oriC* primosome. This requires the stable binding of the *dnaB* protein on a single-stranded sequence at the replication fork through the prior interactions of *dnaA* and *dnaC* at *oriC*. By means of interactions with primase as in the non-specific priming reaction, the *dnaB* protein then directs the formation of RNA primers (LeBowitz and McMacken, 1986; Baker *et al.*, 1986, 1987). Leading strand initiations probably occur immediately following strand opening and are initiated between *dnaA* boxes R2 and R4 in the right-hand part of *oriC* (Seufert and Messer, 1987). Such a primosome is also implicated in the replication of phage λ (Dodson *et al.*, 1986).

The processive movement of the dnaB protein at the replication fork also generates a mobile primosome able to direct the multiple primings of lagging-strand synthesis (Mok and Marians, 1987). This scheme is analogous to the priming systems characterized for the T7 gene 4 protein 63- and 56-kDa forms (Nakai and Richardson, 1988; Figure 6.15) and the phage T4 gene 41 and 61 proteins (Cha and Alberts, 1989; Figure 1.6). Both pairs of proteins form unique isolable complexes that have both helicase and primase activity. These complexes act at the replication fork while bound to the lagging strand template, functioning both to unwind the parental DNA and to synthesize primers that are used to initiate leading strand synthesis at origins and the synthesis of Okazaki fragments on the lagging strand.

Although the *oriC* primosome is capable of initiating lagging strand synthesis, the primosome required for replication of the bacteriophage ϕ X174 single-stranded DNA template (Figure 1.4) has long served as a model for the molecular mechanisms involved in the discontinuous synthesis of the lagging strand (Arai *et al.*, 1981f). Assembly of the primosome is site-specific (Arai and Kornberg, 1981a; Zipursky and Marians, 1981) and requires the action of six separate proteins including PriA, PriB, PriC (formerly proteins *n'*, *n* and *n''*), dnaT (protein *i*), dnaC, and dnaB (Shlomain *et al.*, 1981). Assembly is initiated by PriA protein binding a specific sequence in the SSB-coated single-stranded ϕ X174 viral genome called a primosome assembly site (*pas*). The *pas* sequences from several sources show considerable potential secondary structure (Greenbaum and Marians, 1984, 1985; Masai *et al.*, 1990a). This is probably responsible for their recognition by the PriA protein as these sequences have little sequence similarity or other discernible relationship (Masai *et al.*, 1990a).

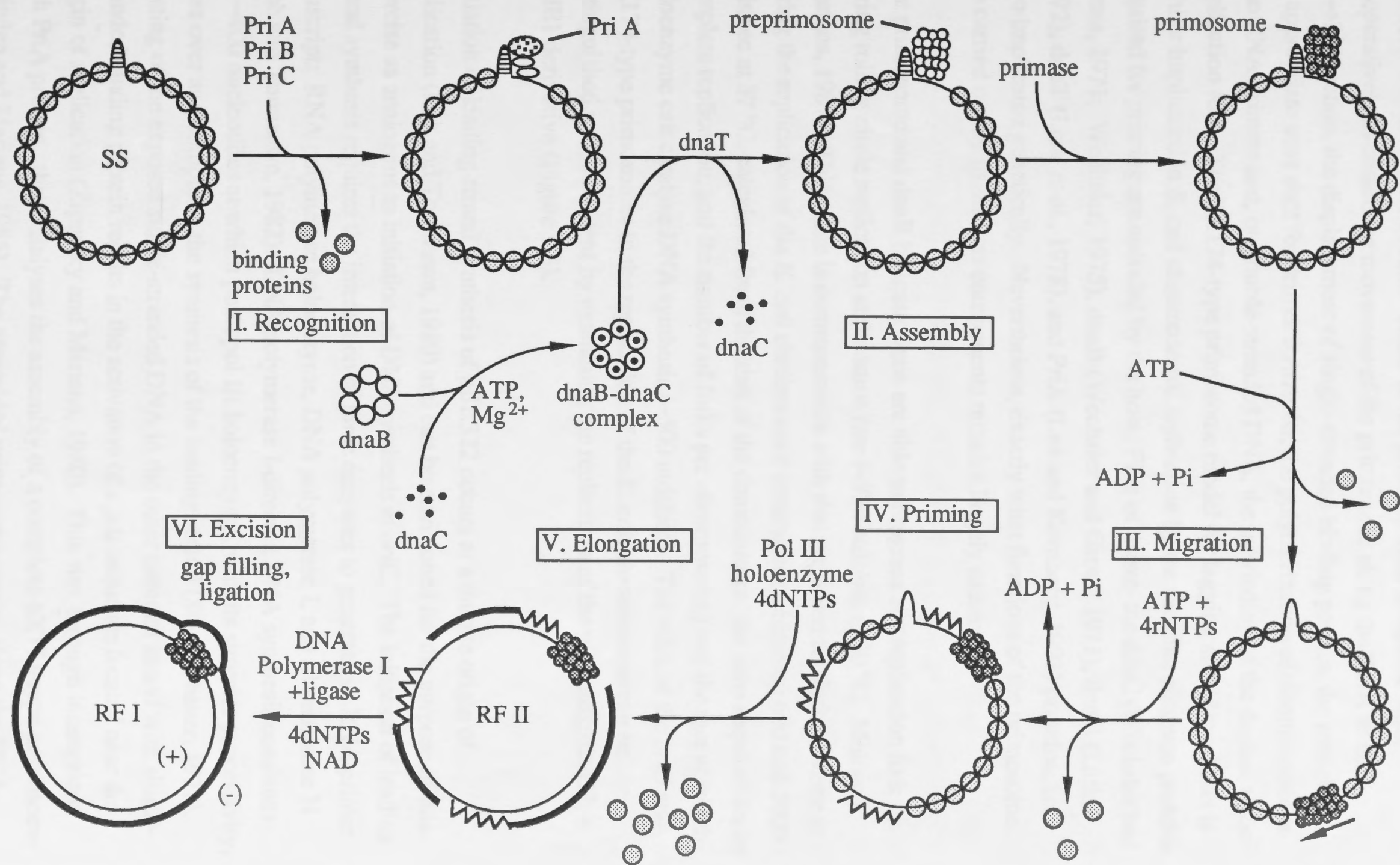
In ϕ X174 replication, four prepriming proteins, PriA, PriB, PriC and dnaT, are required for loading the dnaB protein on the template (Arai *et al.*, 1981f; Figure 1.4). Following recognition and site-specific binding of *pas* by a single molecule of PriA protein (Greenbaum and Marians, 1984; Shlomain and Kornberg, 1980), the PriB, PriC and dnaT proteins may function as a protein bridge that physically links PriA to the dnaB-dnaC complex. The dnaC protein serves a critical role in primosome assembly, but is not maintained in the translocating primosome (Wahle *et al.*, 1989a, b). The interactions between dnaB and dnaC are therefore probably only required at the stage of primosome assembly.

The primosome is processive on the ϕ X174 viral single-stranded DNA template (McMacken *et al.*, 1977). Once formed, it moves processively on the lagging strand of a duplex DNA (Lee and Marians, 1989). Presumably this occurs through the ATP-dependent translocation of the dnaB protein, which also directs the priming and

Figure 1.4

Specific priming by the ϕ X174-type primosome. Replication of the single-stranded ϕ X174 viral genome requires the interaction of five primosomal proteins: PriB, PriC, dnaT, dnaC, and dnaB, following the recognition of an SSB-dependent structure (*pas*) in the DNA template by PriA. A dnaB-dnaC complex, with the aid of the dnaT protein, transfers the dnaB protein to the template forming a preprimosomal assembly. Upon the association of primase with the assembled prepriming proteins the ϕ X174-type primosome is generated. With the release of the dnaC protein the ATP-dependent translocation of dnaB enables the primosome to track along the ϕ X174 template synthesizing primers at numerous places. The action of primase is not essential for either the stability or mobility of the preprimosomal complex and may in fact continually associate and dissociate to perform RNA primer synthesis. Following extension of DNA chains by pol III holoenzyme, primers are excised and gaps are filled by DNA polymerase I, the remaining nicks are sealed by ligase to complete the double-stranded replicative form (RF).

Figure adapted from Kornberg and Baker (1991).



repeated initiation of Okazaki fragments on the lagging strands (Minden and Marians, 1985; Masai and Arai, 1988a). These seven proteins must somehow work cooperatively to ensure the movement of the primosome along the DNA in the 5'→3' direction, the displacement of single-stranded binding protein, the recognition of appropriate start sites for primer formation, the polymerization of ribonucleotides into RNA primers and, on double-stranded DNA, the unwinding of the duplex at the replication fork. The ϕ X174-type primosome model for lagging strand synthesis is further implicated in *E. coli* chromosomal replication by the fact that all seven proteins required for priming are encoded by the host. Five of these; the *dnaC* (Wechsler and Gross, 1971; Wechsler, 1975), *dnaB* (Wechsler and Gross, 1971), *dnaG* (Lark, 1972), *dnaT* (Lark *et al.*, 1978), and *PriA* (Lee and Kornberg, 1991) proteins, are also implicated genetically. Nevertheless, exactly what functions of the primosome are carried out by its various components remains largely unknown.

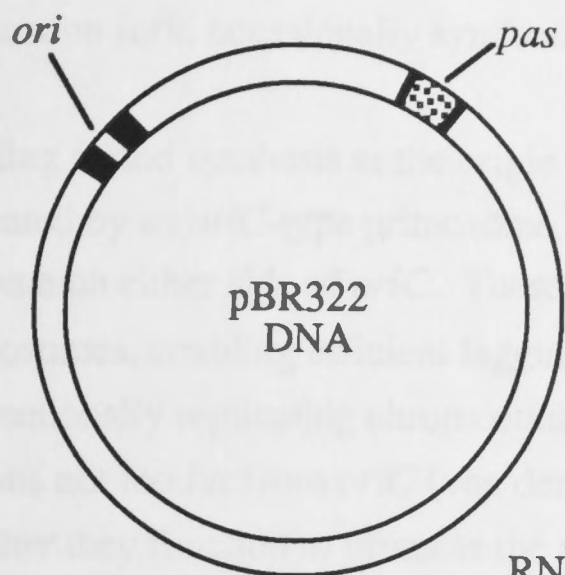
The primosome and *dnaB* helicase alone are able to progress at a replication fork during rolling circle replication at the same rate (~730 ntds/sec at 30 °C; Mok and Marians, 1987). This value is commensurate with that proposed for fork movement during the replication of the *E. coli* chromosome (estimated at between 600 and 1000 ntds/sec at 37 °C, calculated from the size of the chromosome, the time required for its complete replication, and the number of forks per chromosome) and the rate at which holoenzyme can catalyse DNA synthesis (~500 ntds/sec). The roles of the *oriC* and ϕ X174-type primosomes in the replication of the *E. coli* chromosome can be distinguished to some extent by examining the replication of the plasmid pBR322, a ColE1 derivative (Figure 1.5).

Initiation of leading strand synthesis of pBR322 occurs at a unique origin of replication (Itoh and Tomizawa, 1980) and can be envisioned for the purpose of this exercise as analogous to initiation of DNA synthesis at *oriC*. The initiation of leading strand synthesis requires the interaction of three enzymes to generate an RNA primer transcript: RNA polymerase holoenzyme, DNA polymerase I, and ribonuclease H (Itoh and Tomizawa, 1982). DNA polymerase I-directed DNA synthesis continues for ~400 nucleotides at which point pol III holoenzyme, with its superior processivity, takes over and completes the synthesis of the leading strand (Staudenbauer, 1977). Coating of the exposed single-stranded DNA in the other parental strand with single-stranded binding protein results in the activation of a *pas* sequence located near the origin of replication (Zipursky and Marians, 1980). This site, through interactions with *PriA* protein, then catalyses the assembly of a complete ϕ X174-type primosome (Minden and Marians, 1985). The assembled primosome moves along the DNA template in the 5'→3' direction, presumably as a result of the ATP-dependent

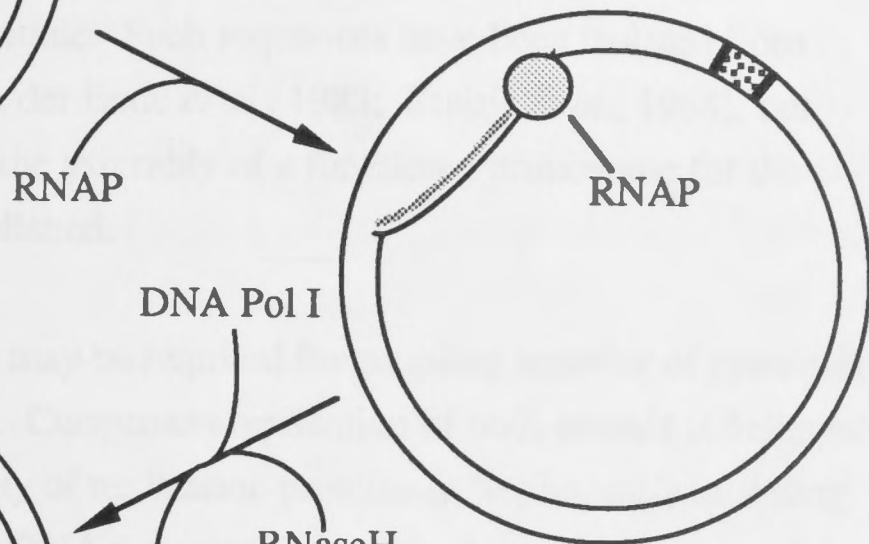
Figure 1.5

A scheme for the initiation of synthesis of both the continuous (leading) and discontinuous (lagging) strands of pBR322. Three host enzymes are required to initiate leading strand synthesis on a negatively supercoiled plasmid *in vitro*: RNA polymerase (RNAP) holoenzyme, DNA polymerase I (DNA Pol I), and ribonuclease H (RNaseH). RNA polymerase holoenzyme synthesizes a primer for leading strand synthesis and ribonuclease H, specific for cleavage of RNA in RNA-DNA hybrids, generates 3'-OH ends in the origin that serve as primers for DNA synthesis by DNA polymerase I. The synthesis of the leading strand by DNA polymerase I probably activates the *pas* sequence on the lagging-strand template by rendering it single-stranded. Lagging-strand synthesis on the pBR322 template then initiates at this *pas* sequence, located 150-bp downstream of the leading-strand origin, a process which requires the same seven proteins essential for priming ϕ X174 replication (Figure 1.4). The primosome, through the actions of the *dnaB* protein, functions as a helicase to unwind the duplex at the replication fork and also initiates priming by primase.

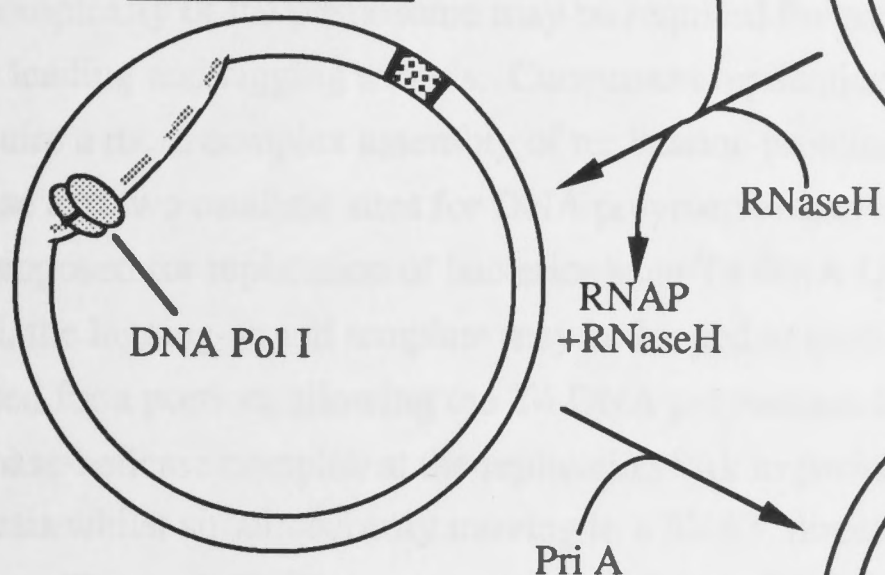
Figure adapted from Kornberg and Baker (1991).



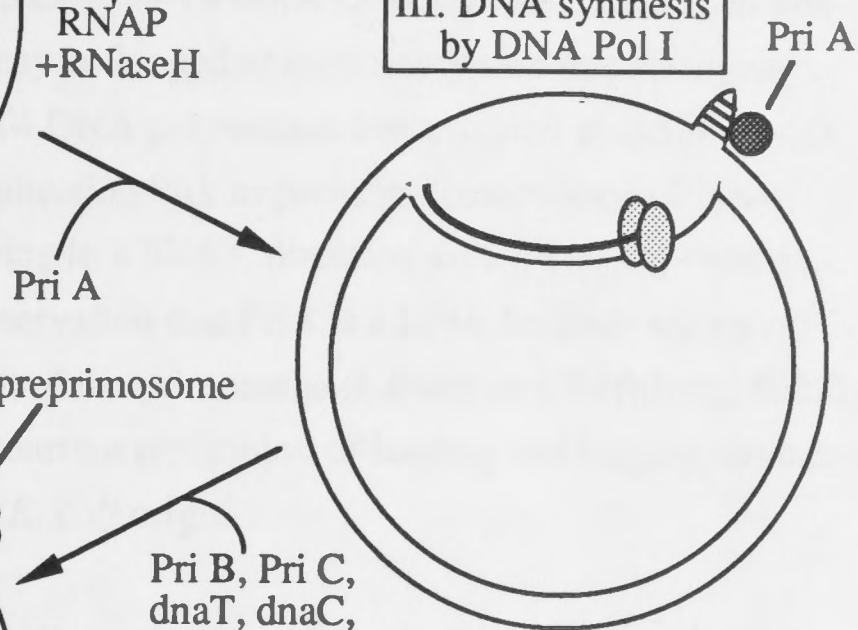
I. Transcription initiation and elongation of hybrid



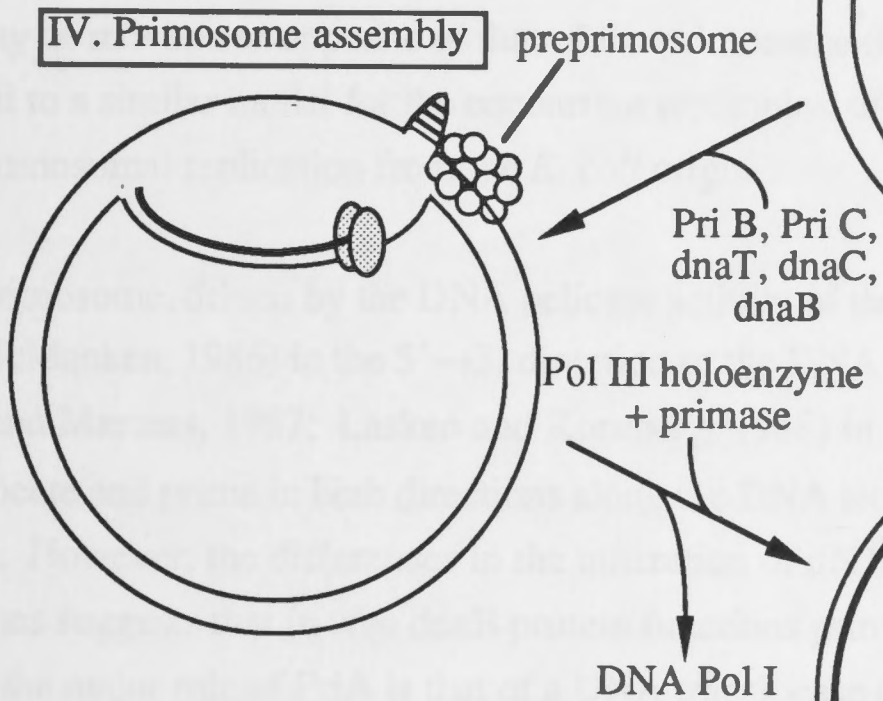
II. Cleavage by RNaseH



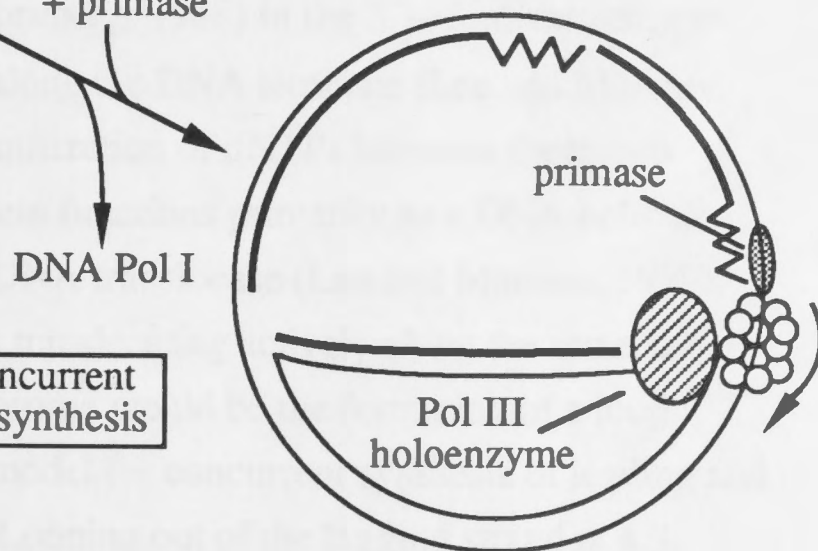
III. DNA synthesis by DNA Pol I



IV. Primosome assembly



V. Concurrent DNA synthesis



translocation of dnaB helicase. The primosome may also unwind the DNA at the replication fork, occasionally synthesizing short RNA primers on the lagging strand.

Leading strand synthesis at the origin of replication of the *E. coli* chromosome, indicated by an *oriC*-type primosome, could likewise activate *pas* sequences in the genome on either side of *oriC*. These may then serve for assembly of ϕ X174-type primosomes, enabling efficient lagging-strand synthesis at both forks of a bidirectionally replicating chromosome. Such sequences have been isolated from regions not too far from *oriC* (van der Ende *et al.*, 1983; Stuitje *et al.*, 1984), but whether they function to promote the assembly of a functional primosome for the lagging strand remains to be established.

The complexity of the primosome may be required for coupling together of synthesis of the leading and lagging strands. Concurrent replication of both strands is believed to require a more complex assembly of replication proteins (a 'replisome') including primase and two catalytic sites for DNA polymerization. Such a model system was first proposed for replication of bacteriophage T4 DNA (Alberts *et al.*, 1982). In this model, the lagging-strand template may be looped in such a way that its polarity is reversed for a portion, allowing the T4 DNA polymerase holoenzyme associated with a primase-helicase complex at the replication fork to perform discontinuous DNA synthesis whilst simultaneously moving in a 5' \rightarrow 3' direction on the lagging-strand template (Figure 1.6). The recent observation that PriA is a DNA helicase with a polarity of movement opposite to that of the primosome (Lasken and Kornberg, 1988) has led to a similar model for the concurrent replication of leading and lagging strands in chromosomal replication from the *E. coli* origin.

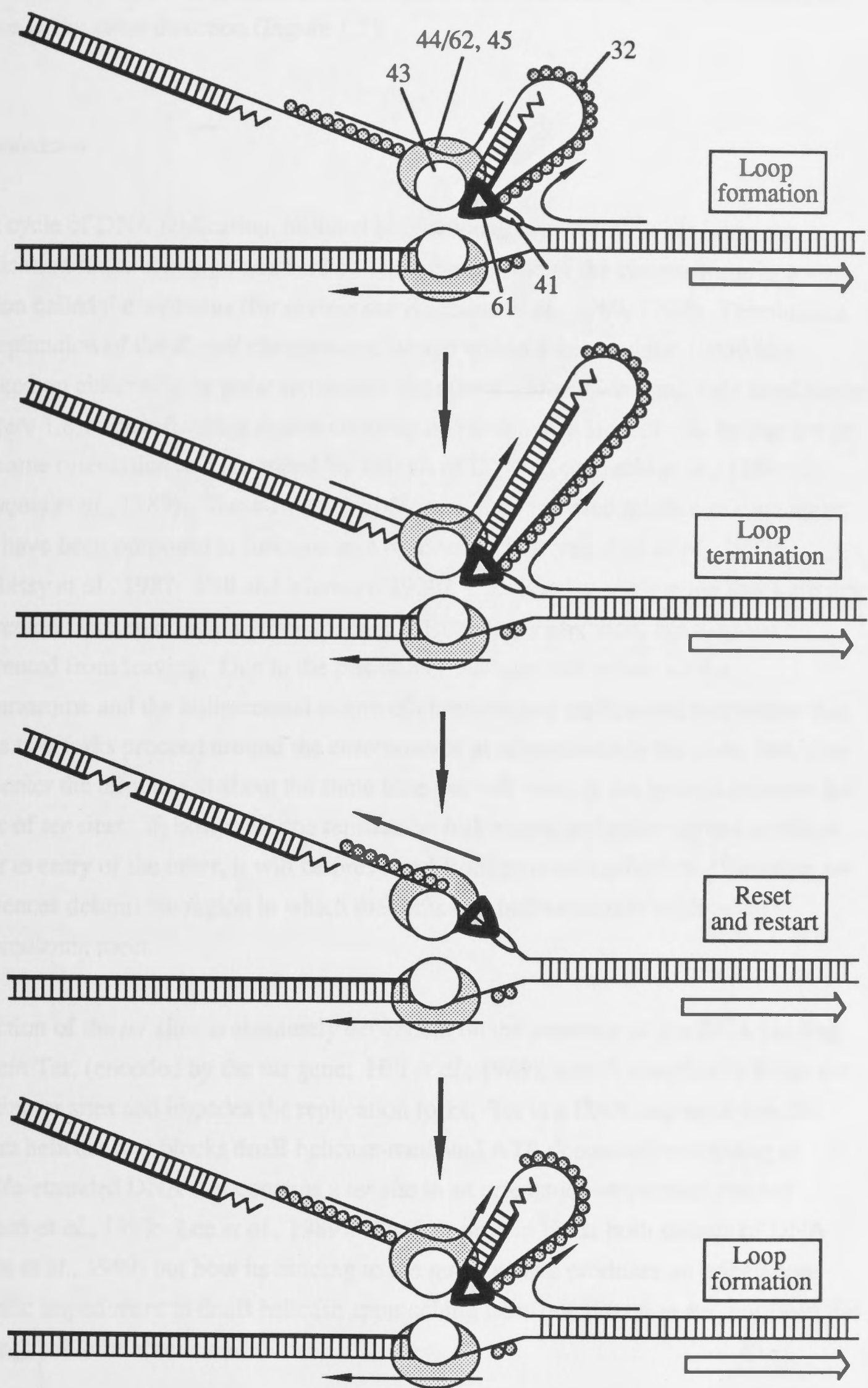
The primosome, driven by the DNA helicase activity of the dnaB protein (LeBowitz and McMacken, 1986) in the 5' \rightarrow 3' direction or the DNA helicase activity of PriA (Lee and Marians, 1987; Lasken and Kornberg, 1988) in the 3' \rightarrow 5' direction, can translocate and prime in both directions along the DNA template (Lee and Marians, 1989). However, the differences in the utilization of dNTPs between these two enzymes suggests that *in vivo* dnaB protein functions primarily as a DNA helicase, while the major role of PriA is that of a DNA translocase (Lee and Marians, 1989). One consequence of both DNA helicases translocating actively along the same DNA strand in opposite directions in the primosome would be the formation of a loop structure similar to that proposed in the model for concurrent synthesis of leading and lagging strands in T4 DNA replication. Looping out of the lagging strand at a chromosomal replication fork reverses its polarity, which could thereby allow the

Figure 1.6

Proposed model of a replication fork in bacteriophage T4. Seven T4-encoded proteins are required for DNA synthesis at a fork *in vitro*. These are indicated by their gene numbers: 43, T4 DNA polymerase; 44, 62, and 45, polymerase accessory proteins; 32, single-stranded binding protein; and 41 and 61, primase-helicase complex. The open arrow indicates the direction of replication fork movement and other arrows indicate the direction DNA moves through the protein complex (Hinton *et al.*, 1987).

T4 DNA polymerase (43) carries out strand displacement synthesis on the leading strand only. It is associated in a holoenzyme with the genes 44/62 and 45 polymerase accessory proteins, which increase the binding of the polymerase to the primer template. Gene 32 DNA binding protein helps to destabilize the DNA duplex. The T4 polymerase holoenzyme complex synthesizing the discontinuous fragments on the lagging strand is held at the fork by binding the 41-61 protein complex. The T4 polymerase selects new priming sites synthesized by the 41-61 helicase-primase complex for chain elongation, a process which could be coupled to completion of synthesis of the previous Okazaki fragment. The 3' → 5' replication of the lagging strand by T4 polymerase and the 5' → 3' translocation of the T4 primosome on the lagging strand leads to the production of a growing loop, half of which is single stranded.

The gene 41 helicase-primase protein utilizes the energy of GTP or ATP hydrolysis to move along the lagging-strand template and unwind the DNA helix ahead of the advancing replication complex. The gene 41 protein also interacts with the gene 61 protein to form the 'T4 primosome'. Together, these two proteins synthesize the pentameric RNA primers that prime the synthesis of Okazaki fragments on the lagging strand. Upon reaching the 3'-OH terminus of a completed Okazaki fragment the polymerase accessory protein complex releases the template. This may act as a signal for the attached primase-helicase complex to make a new primer, which is then recognized and extended by the DNA polymerase. This loop formation and release has been termed the 'trombone model' of lagging-strand DNA synthesis.



dimeric holoenzyme to synthesize nascent DNA on both strands while continuing to move in the same direction (Figure 1.7).

Termination

The cycle of DNA replication, initiated bidirectionally from *oriC*, ends when the replication forks converge and fuse on the opposite side of the chromosome in a region called the terminus (for review see Kuempel *et al.*, 1989, 1990). Termination of replication of the *E. coli* chromosome occurs within a large region (~350 kb) flanked on either side by polar terminator sites (*ter* sites) at its left and right hand limits (Figure 1.8). Each flanking region contains two terminator sites of ~20 bp that are in the same orientation and separated by >50 kb of DNA (Kobayashi *et al.*, 1989; François *et al.*, 1989). These two pairs of terminators inverted relative to each other, and have been purposed to function as a replication fork trap (Hill *et al.*, 1987; deMassy *et al.*, 1987; Hill and Marians, 1990). Specifically, replication forks are not prevented from entering the terminus region from either direction, but they are prevented from leaving. Due to the position of the terminus region on the chromosome and the bidirectional nature of chromosomal replication, this means that if the two forks proceed around the chromosome at approximately the same rate, they will enter the terminus at about the same time and will meet in the interval between the pairs of *ter* sites. If, however, one replication fork enters and traverses the terminus prior to entry of the other, it will be prevented from proceeding further. Therefore *ter* sequences delimit the region in which the forks of a bidirectionally replicating chromosome meet.

Function of the *ter* sites is absolutely dependent on the presence of the DNA-binding protein Ter, (encoded by the *tus* gene; Hill *et al.*, 1989), which specifically binds the terminator sites and impedes the replication forks. Ter is a DNA sequence-specific contra helicase and blocks dnaB helicase-mediated ATP-dependent unwinding of double-stranded DNA that contains a *ter* site in an orientation-dependent manner (Khatri *et al.*, 1989; Lee *et al.*, 1989). The Ter protein binds both strands of DNA (Sista *et al.*, 1989) but how its binding to the *ter* sequence produces an orientation-specific impediment to dnaB helicase approaching from one direction and not from the other is yet to be elucidated.

Following complete replication, circular chromosomes such as that of *E. coli* are extensively intertwined. The complete topological unlinking of the two nascent daughter molecules is required prior to or during cell division (reviewed by Wang,

Figure 1.7

A hypothetical scheme for concurrent replication of leading and lagging strands in replication of the *E. coli* chromosome. The dimeric DNA polymerase III holoenzyme associates with the primosome to form a 'replisome' and anchors the replication fork. In this model DNA synthesis on the lagging strand is accomplished by advancing a loop of the template through this replication complex. Translocation of PriA on the lagging strand template with 3'→5' polarity, coupled to translocation of the dnaB protein as it unwinds the duplex in a 5'→3' direction could function together to feed the DNA loop. The thick bars represent RNA primers synthesized by primase in response to secondary structures engineered in the single-stranded template by the dnaB protein and arrows indicate the direction DNA moves through the protein complex. The open arrow indicates the direction of fork movement.

Figure adapted from Lasken and Kornberg (1988).

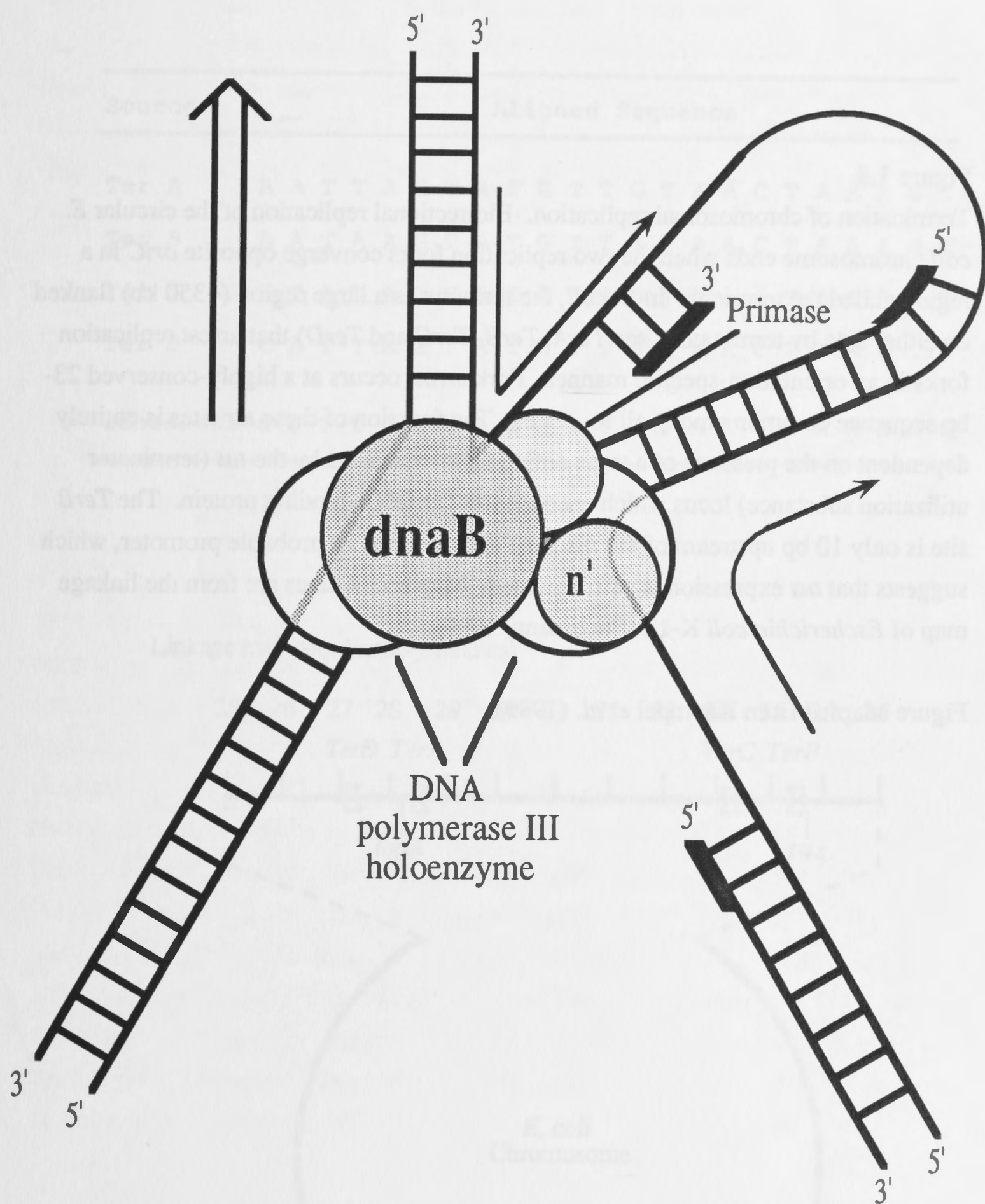


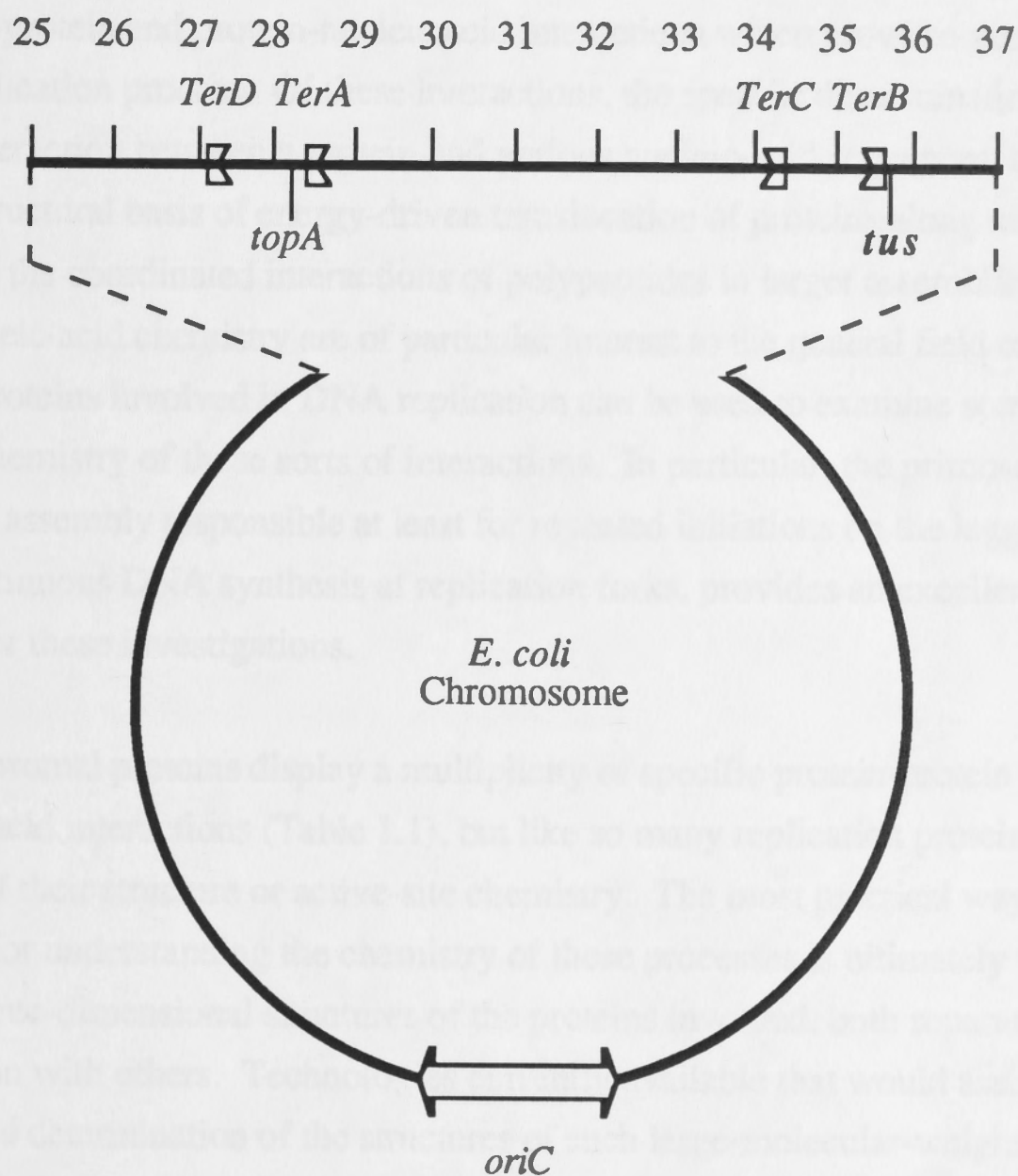
Figure 1.8

Termination of chromosomal replication. Bidirectional replication of the circular *E. coli* chromosome ends when the two replication forks converge opposite *oriC* in a region called the terminus. In *E. coli*, the terminus is a large region (~350 kb) flanked on either side by terminator sites (*TerA*, *TerB*, *TerC* and *TerD*) that arrest replication forks in an orientation-specific manner. Fork arrest occurs at a highly-conserved 23-bp sequence common among all four sites. The function of these *ter* sites is entirely dependent on the presence of a trans-acting factor specified by the *tus* (terminator utilization substance) locus which encodes the Ter DNA-binding protein. The *TerB* site is only 10 bp upstream of the *tus* gene and overlaps its probable promoter, which suggests that *tus* expression is autoregulated. Map coordinates are from the linkage map of *Escherichia coli* K-12 (Bachmann, 1990).

Figure adapted from Kuempel *et al.* (1989).

Source	Aligned Sequence
<i>Ter A</i>	A A T T A G T A T G T T G T A A C T A A A G T
<i>Ter B</i>	A A T A A G T A T G T T G T A A C T A A A G T
<i>Ter C</i>	A T A T A G G A T G T T G T A A C T A A T A T
<i>Ter D</i>	C A T T A G T A T G T T G T A A C T A A A T G
Consensus	a a t t A G t A T G T T G T A A C T A A a n t

Linkage map coordinates (minutes)



1991). The action of gyrase, which is capable of transiently breaking phosphodiester bonds on both strands, enabling passage of one strand through the other (Mizuuchi *et al.*, 1980), decatenates the daughter chromosomes to enable chromosome segregation and cell division to proceed normally (Steck and Drlica, 1984). Topoisomerase I, which functions by a single-strand passage mechanism, may also decatenate duplex circles if at least one of the two members of the catenane contains a nick or gap (DiGate and Marians, 1988). Gyrase also plays an important role in relieving superhelical tension ahead of a fork, and may be able to control the rate a fork approaches the terminus by regulating the degree of superhelical tension in the DNA.

Scope of the thesis

It is evident that research involving proteins essential for the replication of the *E. coli* chromosome has contributed in part to the rapid expansion in knowledge of the nature of interactions between proteins and nucleic acids. The functions of many of these proteins that promote the replication of DNA are now well understood. However, there is essentially no knowledge of their active-site chemistry or the chemistry of the specific protein-protein and protein-nucleic acid interactions which prove so vital to a coordinated replication process. Of these interactions, the specific discrimination and chemistry of interaction between a protein and various nucleic-acid sequences, the chemical and structural basis of energy-driven translocation of proteins along nucleic-acid chains, and the coordinated interactions of polypeptides in larger assemblies involved in nucleic-acid chemistry are of particular interest to the general field of biochemistry. Proteins involved in DNA replication can be used to examine some aspects of the chemistry of these sorts of interactions. In particular, the primosome, the multiprotein assembly responsible at least for repeated initiations on the lagging strand in discontinuous DNA synthesis at replication forks, provides an excellent model system for these investigations.

The seven primosomal proteins display a multiplicity of specific protein-protein and protein-nucleic acid interactions (Table 1.1), but like so many replication proteins very little is known of their structure or active-site chemistry. The most practical way to provide a basis for understanding the chemistry of these processes is ultimately to determine the three-dimensional structures of the proteins involved, both separately and in association with others. Technologies currently available that would assist or allow the detailed determination of the structures of such large-molecular-weight assemblies and their components are X-ray crystallography, high-resolution electron microscopy of two-dimensional crystals, and high-field nuclear magnetic resonance

Table 1.1 The primosomal proteins of *Escherichia coli*.

Protein	Encoding gene	Map position (minutes)	Estimated molecules per cell	Subunit molecular weight (M_r)	Subunits in native protein	Functions
dnaT (protein i)	<i>dnaT</i>	99	50	19,500	3	Primosome assembly and possibly other primosomal functions as yet unknown.
dnaC(D)	<i>dnaC</i>	99	100	27,900	1	Forms an ATP-dependent dnaB-dnaC (6:6) complex. Facilitates interactions between single-stranded DNA and dnaB protein. Essential for loading dnaB protein onto SSB-coated single-stranded DNA through interactions with dnaA or other primosomal proteins.
dnaB	<i>dnaB</i>	92	20	52,300	6	ATP-dependent generation of primase recognition sites. 5'→3' DNA helicase activity involved in the <i>oriC</i> opening reaction and primosome translocation.
PriA (protein n')	<i>priA</i>	88	70	81,700	1	ATP-dependent recognition of secondary structural features in DNA leading to primosome assembly. 3'→5' DNA helicase activity involved in either leading-strand primosome translocation or translocation on the lagging-strand template during concurrent template replication by the replisome.
PriB (protein n)	<i>priB</i>	96	80	11,400	2	Primosome assembly and possibly other primosomal functions as yet unknown.
PriC (protein n'')	<i>priC</i>	11	?	20,300	1	Primosome assembly and possibly other primosomal functions as yet unknown.
dnaG (primase)	<i>dnaG</i>	67	50	65,500	1	Primer synthesis in response to secondary structural elements in the DNA template.

spectroscopy (NMR). Each of these tools requires the availability of large stocks of protein. Unfortunately, like many of the enzymes of *E. coli* DNA replication, the primosomal proteins are present in normal cells at very low levels (Table 1.1).

The low level of natural production of the proteins of the primosome is a result of severe regulation at the levels of transcription of the genes and translation of the mRNA and probably reflects the high efficiency of their actions and the fact that they are required only once in the lifetime of the cell. Initial goals were therefore directed at obtaining richer sources of primosomal proteins whose genes were currently available, to enable their isolation in quantities sufficient for chemical and structural studies. For the specific purpose of overproducing many of the replication proteins of *E. coli* for similar studies, members of this laboratory have developed a modified new series of bacteriophage λ promoter vectors (Elvin *et al.*, 1990). By use of these vectors and modern methods in recombinant DNA technology, it was possible to overcome the restrictions of natural gene regulation and to overproduce some of the proteins of the *E. coli* primosome.

Procedures for large-scale isolation of the proteins have been devised. They have been characterized in several ways, and the stage is now set for studies that should give answers to questions like those raised above.

BACTERIAL STRAINS AND PLASMID VECTORS

2.1.1 Bacterial Strains

The strains of *Escherichia coli* used in these studies are listed in Table 2.1.

2.1.2 Plasmid Vectors

The plasmid vectors used in these studies are listed in Figure 2.1, and are described in detail in the legend. Vectors pCE30, pCE33 and derivatives have been described (Elvin *et al.*, 1990). The series of pMTL-P vectors (Chambers *et al.*, 1988) were generously provided by Dr N.P. Minton (PHLS Centre for Applied Microbiology and Research, Wiltshire).

2.1.3 Growth Media

E. coli strains were grown in LB medium (Luria and Burrous, 1957) supplemented with 25 µg/ml thymine (LBT) and, as required, with ampicillin (50 µg/ml) and/or kanamycin (50 µg/ml) and/or chloramphenicol (7 µg/ml). Strains containing pCE30 and its derivatives were grown at 30 °C, others at 37 °C. Strains prepared for large scale plasmid preparations were grown in minimal 56 medium (Monod *et al.*, 1951) containing 1 mM magnesium chloride and supplemented with trace metals (Gibson *et al.*, 1977), 40 mM glucose, 1 µg/ml (w/v) vitamin B1 and 50 µg/ml ampicillin.

CHEMICALS, REAGENTS AND INSTRUMENTS

All chemicals and reagents listed in the text, unless otherwise stated, were analytical grade and were obtained from BDH Chemicals, Ajax Chemicals, Mallinckrodt Chemical Works, May and Baker, Calbiochem, Fluka AG, Cabot or the Sigma Chemical Company. Acrylamide, bis-acrylamide and the antibiotics ampicillin, chloramphenicol and kanamycin were obtained from Boehringer-Mannheim.

Table 2.1

Escherichia coli strains and relevant genotypes.

Strain	Genotype	Source/Reference
AN1459	<i>supE44 thi-1 leuB6 thr-1 ilvC hsdR</i> <i>recA srlA::Tn10</i>	Elvin <i>et al.</i> (1986)
AN2666 (JM101 <i>recA</i>)	<i>supE44 thi-1 Δ(lac-proAB) recA</i> <i>srlA::Tn10</i>	Laboratory stock (Yanisch-Perron <i>et al.</i> , 1985)
K37	<i>HfrC supD32 relA1 pit-10 spoT1</i> <i>tonA22 ompF627 phoA4 λ⁻</i>	Hines and Ray (1980)
RSC453 (SG1692 <i>recA</i>)	<i>F⁻ thi leu his argG metB thyA deo</i> <i>lacY malA xyl-7 gal mtl rpsL tonA tsx</i> <i>supE44 dnaB^{ts} λ^r λ⁻ recA srlA::Tn10</i>	Laboratory stock (Sevastopoulos <i>et al.</i> , 1977)
AR120	<i>λcI⁺ Δ-gal nadA::Tn10</i>	Mott <i>et al.</i> (1985)

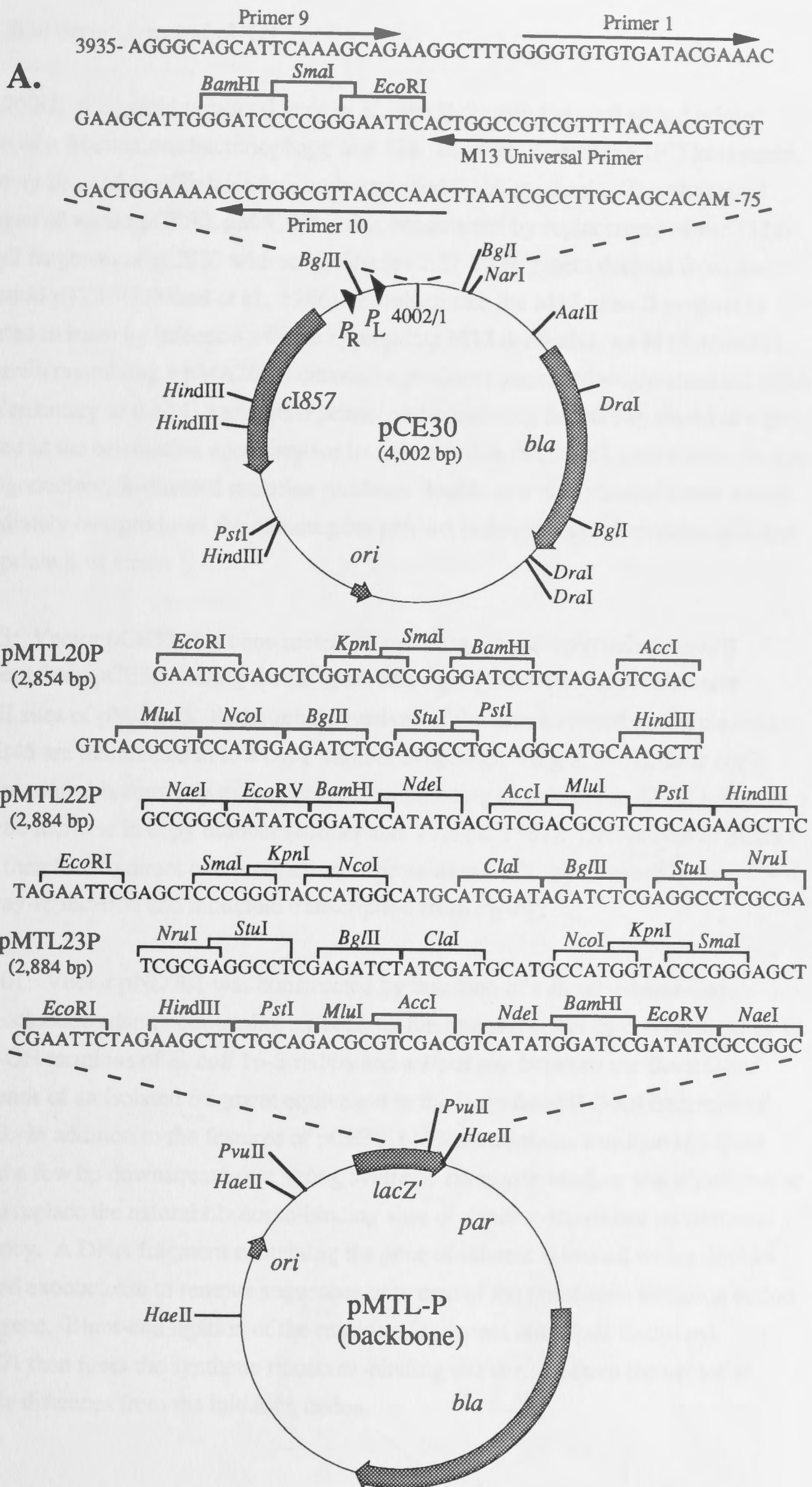
Figure 2.1

A. pCE30 and the pMTL-P vectors

pCE30: The work in this thesis required the extensive use of the vector pCE30 and its derivatives (Elvin *et al.*, 1990). The expression vector pCE30 was engineered by inserting a 1349-bp *Pst*I/*Bam*HI DNA fragment bearing the *cI*857^{ts} gene and the two λ promoters between the *Pst*I and *Bam*HI sites in the polylinker region of pUC9. This construction placed the tandem λ promoters immediately upstream of unique *Bam*HI, *Sma*I and *Eco*RI restriction endonuclease cleavage sites suitable for insertion of genes to be expressed. In strains growing at 30 °C, a pCE30 derivative directs synthesis of the *cI*^{ts} repressor, which prevents transcription from the λ promoters. Upon heating of cultures to 42 °C, the thermolabile repressor is inactivated, and high-level expression of genes downstream of the promoters is enabled. Since the *cI*857 gene is expressed by the plasmid itself, pCE30 derivatives may be used in virtually any *E. coli* strain. By analogy with pUC9, their copy numbers are expected to be about 500 per cell (Chambers *et al.*, 1988). This simplifies isolation and manipulation of plasmid DNA, and the high gene dosage might be expected to contribute to the levels of gene expression obtainable. It should be noted, however, that construction of pCE30 interrupted the reading frame of *lacZ'* in pUC9, so that it is no longer possible to screen directly by colour for recombinant products of a cloning experiment.

Displayed is the genetic and physical map of vector pCE30 (4002 bp), showing the locations of and direction of transcription from the tandem bacteriophage λ promoters *P_R* and *P_L*, and the *bla* (ampicillin resistance gene) and *λcI*857 genes. Relevant restriction endonuclease sites are indicated. The expanded DNA sequence shows the location of the unique *Bam*HI, *Sma*I and *Eco*RI sites useful for insertion of genes, together with the actual sequence (primers 1 and 9) or complements of sequences (M13 universal primer and primer 10) of 17-mer oligonucleotide primers used for DNA sequencing.

pMTL-P plasmids: The second series of cloning vectors used in this study were the pMTL vectors (Chambers *et al.*, 1988). These are similar to the pUC plasmids but carry a more extensive polylinker region within the *lacZ'* gene, and also the *par* region of pSC101 that confers stable high copy numbers even in the absence of ampicillin. The remainder of these vectors was derived from pBR322. The expanded DNA sequences from the three pMTL vectors used in this work show the location of some of the restriction endonuclease sites useful for insertion of genes.



B. pCE30 derivatives and pUC9

pMA200U: Phagemid vectors (Dente *et al.*, 1983) contain the viral strand origin region of a filamentous bacteriophage (*e.g.* f1). In an M13-sensitive (F^+) host strain, they may be used as templates for single-stranded DNA synthesis. The phagemid analogue of vector pCE30, pMA200U, was constructed by replacement of the 1118-bp *Bgl*II fragment of pCE30 with an analogous 1.27-kb fragment derived from the phagemid pTZ18U (Mead *et al.*, 1986). Provided that the M13 gene II product is provided *in trans* by infection with an appropriate M13 derivative, an M13-sensitive host strain containing a pMA200U derivative produces packaged single-stranded DNA complementary to the M13 universal primer and containing the coding strand of a gene inserted in the orientation necessary for its transcription from the λ promoters. Its use for oligonucleotide-directed mutation produces double-stranded plasmid DNA which immediately overproduces the mutant gene product following transformation into any appropriate host strain.

pCE33: Vector pCE33 was constructed by inserting a 1357-bp *Hind*III/*Bam*HI fragment from pCE30 bearing the cI^{ts} gene and P_R - P_L between the *Hind*III and *Bam*HI sites of pMOB45. Although derivatives of the runaway-replication plasmid pMOB45 are maintained at low copy number in cells growing at 30 °C, their copy number control is sensitive to temperature, and heating of cultures to 42 °C results in a dramatic increase in copy number (Bittner and Vapnek, 1981). Derivatives of pCE33 ought therefore to direct overproduction of proteins at 42 °C by virtue of both runaway-replication and inducible transcription from P_R - P_L .

pND201: Vector pND201 was constructed by insertion of a *Bam*HI-blunt-end oligonucleotide adaptor containing a ribosome-binding site perfectly complementary to the 3'-OH terminus of *E. coli* 16-S rRNA and a *Hpa*I site between the *Bam*HI and blunt ends of an isolated fragment equivalent to the large *Bam*HI-*Sma*I fragment of pCE30. In addition to the features of pCE30, pND201 contains a unique *Hpa*I site located a few bp downstream of a strong synthetic ribosome-binding site which can be used to replace the natural ribosome-binding sites of genes to maximize translational efficiency. A DNA fragment containing the gene of interest is treated with a double-stranded exonuclease to remove sequences upstream of the translation initiation codon of the gene. Blunt-end ligation of the resulting fragments into *Hpa*I-linearized pND201 then fuses the synthetic ribosome-binding site derived from the vector at variable distances from the initiating codon.

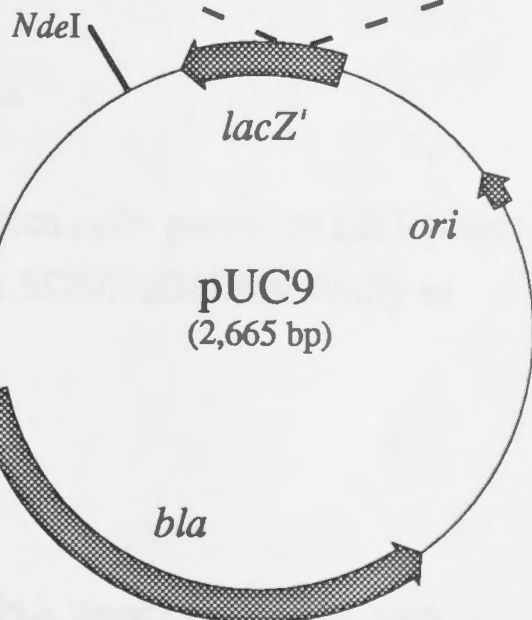
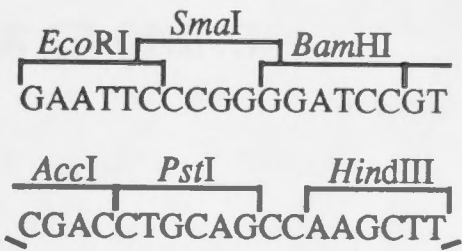
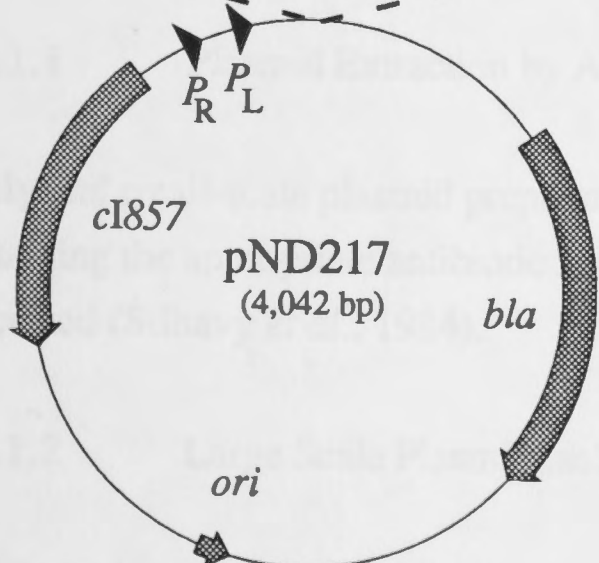
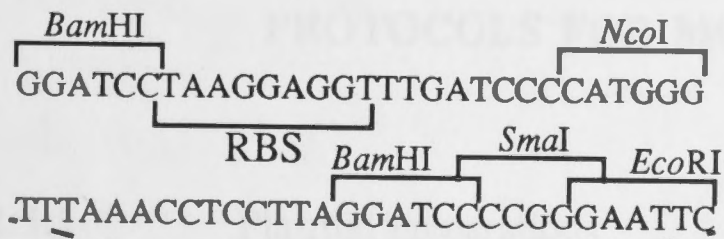
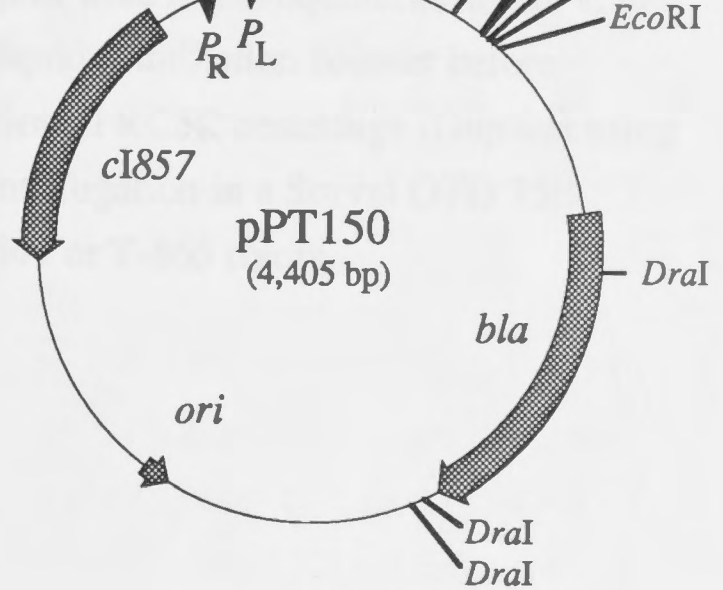
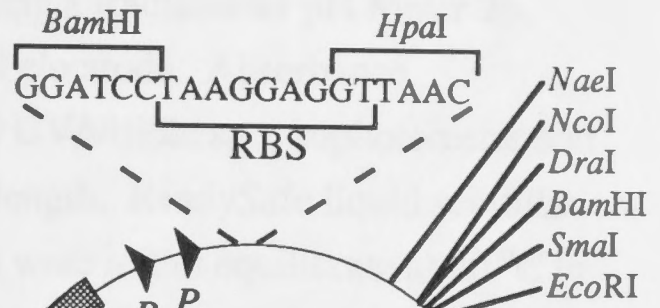
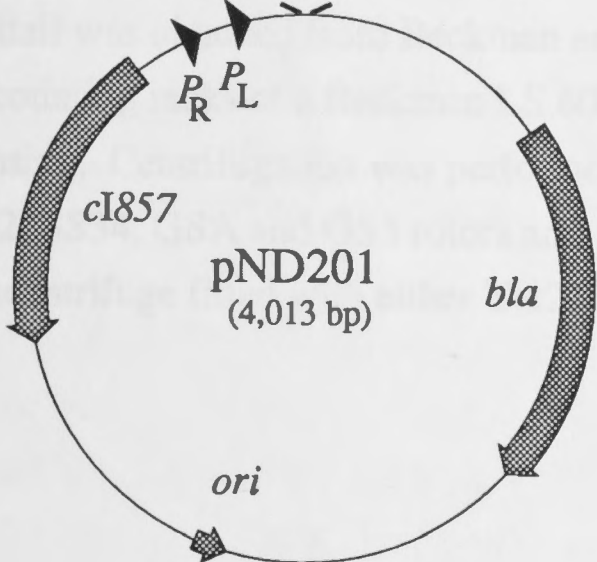
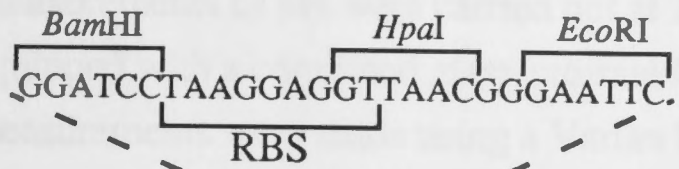
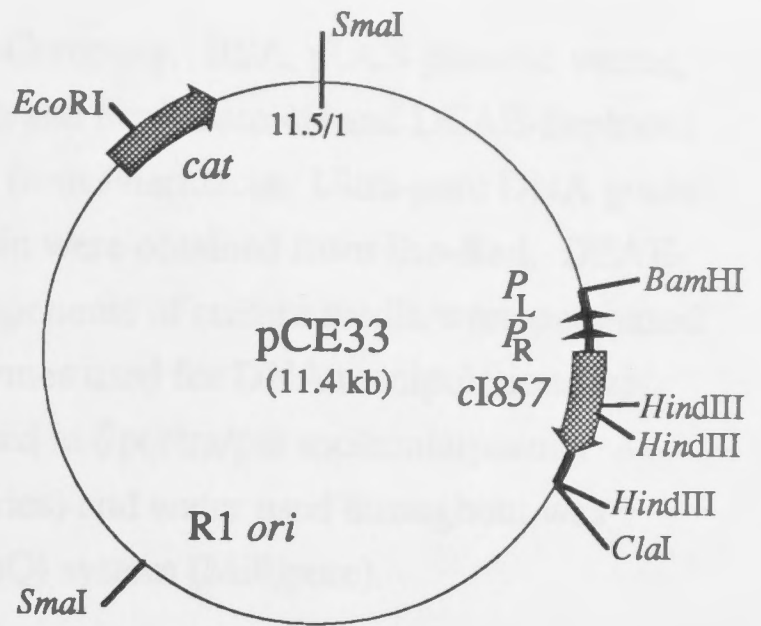
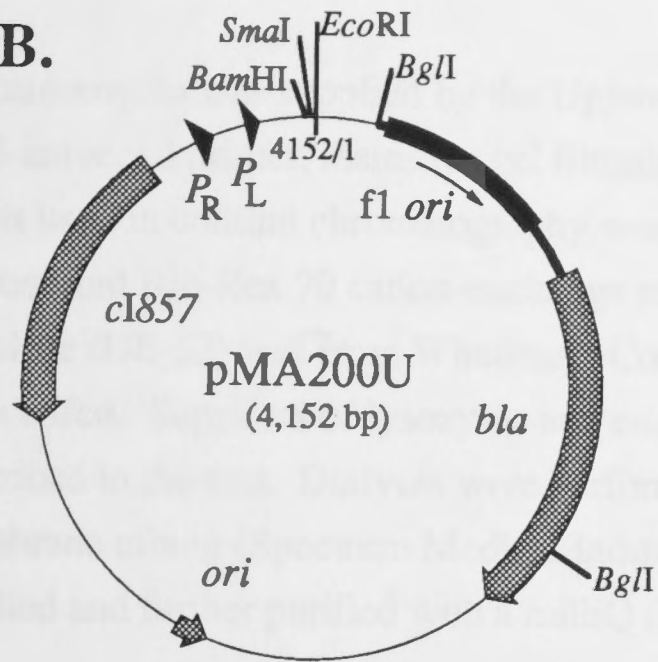
pPT150: pPT150, an analogue of pND201, contains a 363-bp segment derived from pBR322 between the *Hpa*I site and a cluster of restriction endonuclease sites suitable for insertion of a gene. Use of pPT150 for replacement of the natural ribosome-binding site of a gene involves first of all insertion of the gene, with up to ~350 bp of DNA upstream of its translation initiation codon, into the cluster of restriction sites in the orientation necessary for its transcription from P_R - P_L . The isolated product is then linearized at an appropriate site (*e.g.* see Chapter 4) and a double-stranded exonuclease used to remove DNA (from both ends) to the region just upstream of the initiating codon. The product is then cleaved at the unique *Hpa*I site immediately downstream of the synthetic ribosome-binding site. The large fragment is then isolated and recircularized. The advantages of the use of pPT150 over pND201 are that intramolecular recircularization is a much more efficient process than blunt-end ligation of two fragments, and the gene must remain in the required orientation for expression.

pND217: The vector pND217 belongs to a series of seven vectors which have a 9-bp ribosome-binding site perfectly complementary to the 3' end of *E. coli* 16-S rRNA variably spaced by means of an AT rich region from a unique *Nco*I site.¹ Linearization of these vectors with *Nco*I and end-filling with the large fragment of DNA polymerase I results in blunt ends with the sequence 5'.....CCATG^{-OH}. The terminal ATG can then be employed as a start codon for coding regions of genes inserted in-frame between the blunt ends. This strategy is especially useful in generating NH₂-terminal deletions of protein products, or in expressing parts of genes for production of potential antigens.

pUC9: The pUC plasmids were designed for cloning, sequencing and expressing foreign genes in *E. coli* (Vieira and Messing, 1982; Yanisch-Perron *et al.*, 1985). The vector pUC9 consists of the pBR322-derived ampicillin resistance gene (*bla*) and origin of replication, the lac promoter-operator, and a portion of the *lacZ* gene of *E. coli*. A DNA insert containing an array of unique restriction enzyme recognition sites has also been introduced into the *lacZ'* (α -complementing) region enabling cloning of DNA fragments.

1. N.E. Dixon, unpublished.

B.



Spectinomycin was supplied by the Upjohn Company. BSA, pUC9 plasmid vector, M13 universal primer, resins for gel filtration and Sepharose 4B and DEAE-Sephacel resins used in column chromatography were from Pharmacia. Ultra-pure DNA grade agarose and Bio-Rex 70 cation-exchange resin were obtained from Bio-Rad. DEAE-cellulose (DE-52) was from Whatman. Components of culture media were purchased from Difco. Suppliers of lysozyme and enzymes used for DNA manipulations are described in the text. Dialyses were performed in Spectra/por molecularporous membrane tubing (Spectrum Medical Industries) and water used throughout was distilled and further purified with a milliQ (mQ) system (Millipore).

Measurements of pH were carried out at 25 °C using a Radiometer pH Meter 26, equipped with a combined glass-saturated calomel electrode. Absorbance measurements were made using a Varian DMS 90 UV/visible spectrophotometer and with quartz cuvettes (1 ml capacity) of 1 cm path length. ReadySafe liquid scintillation cocktail was obtained from Beckman and samples were left to equilibrate at 20 °C in the counting racks of a Beckman LS 6000 IC liquid scintillation counter before counting. Centrifugation was performed in a Sorval RC5C centrifuge (Dupont) using SE12, SS34, GSA and GS3 rotors and ultracentrifugation in a Sorval OTD 75B ultracentrifuge fitted with either T-1270, TH-641 or T-865 rotors.

PROTOCOLS FOR MOLECULAR CLONING

2.3.1 Plasmid Preparations

2.3.1.1 Plasmid Extraction by Alkaline Lysis

Analytical small-scale plasmid preparations were from cells grown on LBT plates containing the appropriate antibiotic and lysed with SDS/NaOH essentially as described (Silhavy *et al.*, 1984).

2.3.1.2 Large Scale Plasmid Isolation

Highly purified large-scale isolations of plasmid DNA were prepared by two successive bandings in CsCl-density gradients from cells grown in minimal media,

amplified with spectinomycin and lysed with Triton X-100 using a procedure modified from that of Davis *et al.* (1980). All procedures were carried out at 0 °C unless otherwise indicated.

An overnight culture of the plasmid-containing *E. coli* strain was used to inoculate 1-l of 56-minimal medium containing ampicillin. The culture was aerated at 30 °C until $A_{595} = 0.5$. Spectinomycin (300 mg) was added, and aeration was continued at 30 °C for a further 16 h. The cells were chilled and harvested by centrifugation (8,000 x g, 20 min). The cell pellet was resuspended with 6.25 ml of ice-cold 50 mM Tris.HCl pH 8.0, 25 % (w/v) sucrose (resuspension buffer) and stored at -70 °C until required.

The cell suspension was thawed, then diluted to 7.5 ml with resuspension buffer containing lysozyme (Sigma Chemical Company) at 10 mg/ml and swirled for 5 min. To this 1.25 ml of 500 mM EDTA (pH 8.5) was added and the suspension was swirled for a further 5 min. Following the method of Katz *et al.* (1973), cell lysis was performed with the addition of Triton X-100 solution (10 ml; 0.1 % (v/v) Triton X-100, 50 mM Tris.HCl pH 8.0 and 62.5 mM EDTA). After 10 min the lysate was centrifuged (48,000 x g, 1 h) and the clear supernatant was collected by decantation.

The density of the lysis supernatant was adjusted by dilution to 25 ml with mQH₂O and the addition of 24.38 g CsCl. Insoluble material was removed by centrifugation (13,000 x g, 1 h) and the supernatant was decanted through a tissue filter into a Sorvall T-865 polyallomer tube. The tube was filled with the addition of 2.55 ml of a 10 mg/ml solution of ethidium bromide and TE (10 mM Tris. HCl pH 7.4, 1 mM EDTA) + 0.975 g/ml CsCl. The plasmid DNA was isolated in the density gradient produced by centrifugation (125,000 x g, >40 h). The plasmid DNA band, visualised using a longwave UV lamp, was collected from tubes using an 18-gauge hypodermic needle and syringe as described (Sambrook *et al.*, 1989). Plasmid DNA was transferred to a T-1270 polyallomer tube and topped up with 0.88 ml of 10 mg/ml ethidium bromide solution and TE + 0.975 g/ml CsCl. This was centrifuged (125,000 x g, >40 h); the plasmid DNA was isolated as before. Ethidium bromide was removed by repeated extraction with an equal volume of propan-2-ol saturated with 5 M NaCl in TE (10 mM Tris.HCl pH 7.4, 1 mM EDTA). The plasmid solution was then dialysed against 2 changes of 1-l of TE over 12 h.

The DNA concentration was determined spectrophotometrically, assuming a solution with $A_{260} = 1$ contained 50 µg/ml DNA. Plasmids were routinely stored at -70 °C.

2.3.2 Restriction Endonuclease Digestion of DNA

All restriction endonucleases were obtained from New England Biolabs, Amersham International or Boehringer Mannheim. Restriction endonuclease digests were routinely performed in buffers supplied by Boehringer Mannheim wherever possible. If multiple digestions required separate buffer systems, the digestion requiring the lower salt concentration was performed first and subsequent digestions requiring higher salt concentrations were performed after the salt concentration had been elevated to the required level. Digestions were carried out at 37 °C for 1 h and were terminated by the addition of one-half volume of restriction endonuclease stop mix (RE Stop mix: 50 mM EDTA pH 8.5, 17 % (v/v) glycerol and 0.07 % (w/v) bromophenol blue) at 0 °C .

It was occasionally necessary to linearize plasmid DNA containing more than one particular restriction site. To find conditions optimal for production of only a single cleavage of circular plasmid DNA, a series of two-fold dilutions were made of the restriction enzyme. This was done in equivalent volumes containing identical concentrations of buffer and plasmid DNA. These mixtures were treated at 37 °C for 30 min and the reactions were stopped by the addition of 1 volume of RE Stop mix at 0 °C. After visualizing the cleavage patterns following electrophoresis and fragment separation on an appropriate agarose gel (Section 2.3.6), those conditions that gave maximal recovery of linear plasmid DNA were used in a scaled-up partial digestion.

2.3.3 DNA End-filling with the Large Fragment of DNA Polymerase I

Recessed 3'-OH termini created by digestion of DNA with restriction endonucleases were filled-in using the *E. coli* DNA polymerase I Klenow fragment (Klenow enzyme). End-filling reactions were carried out as described (Maniatis *et al.*, 1982) using Klenow fragment supplied by Boehringer Mannheim.

2.3.4 5'-Dephosphorylation of Linearized Plasmid DNA

The 5' phosphate group of linearized plasmid DNA produced by restriction endonuclease digests of plasmid vectors was removed prior to insertions of DNA fragments to prevent intramolecular recircularization of vector fragments. Dephosphorylation reactions were carried out using calf intestinal alkaline phosphatase (Pharmacia) as described (Maniatis *et al.*, 1982).

2.3.5 Bal31 Exonuclease Digestion of Linear DNA

Exonuclease Bal31 is a Ca^{2+} -dependent exonuclease that catalyses the removal of nucleotides from both 5'-P and 3'-OH termini of double-stranded DNA. Activity of Bal31 is inhibited by EDTA which removes calcium from the reaction. To find conditions for the removal of the required number of nucleotides, a trial Bal31 digest was necessary.

A 55 μl reaction volume containing Bal31 buffer (Maniatis *et al.*, 1982) and a known quantity ($\sim 2 \mu\text{g}$) of purified linear DNA was prewarmed at 30 °C for 5 min. Following removal of 5 μl to a tube containing 3 volumes of RE Stop mix at 0 °C, a precise amount of Bal31 (Boehringer Mannheim) was added and mixed. Incubation at 30 °C was continued while transferring 5 μl of the reaction mix into 3 volumes of RE Stop mix at 0 °C every 60 sec. The size range of DNA fragments from each sample was then calculated from electrophoretic mobility on an agarose gel, relative to those of fragment size standards. The conditions that gave DNA fragments of the required size were then used in a scaled-up Bal31 digest.

2.3.6 Agarose Gel Electrophoresis

For both preparative and analytical agarose gel electrophoresis, agarose gels were cast in a Davis system horizontal submarine apparatus (Maniatis *et al.*, 1982) using a toothed comb to form sample wells. Concentrations of 0.7 % to 2.5 % (w/v) agarose for 200 ml standard gels (147 x 136 x 10 mm) and 37.5 ml minigels (75 x 50 x 10 mm) were made to volume with TBE buffer (89 mM Tris-borate, 89 mM boric acid, 2 mM EDTA) (Sambrook *et al.*, 1989) containing 0.2 $\mu\text{g/ml}$ ethidium bromide. Samples were loaded and electrophoresed at 50 V using a Bio-Rad 2000/02 power supply until the DNA had entered the gel matrix. Electrophoresis was then continued at 100 V for the time required for resolution of DNA fragments.

DNA fragments could be visualised with a hand held UV lamp or on a short-wave transilluminator (UV transilluminator Model TS-15, Ultraviolet Products).

Photography was carried out on the transilluminator using a Model MP-4 Land Camera (Polaroid) with Polaroid 4x5 type 57/high speed film. Films were developed according to the manufacturer's instructions.

DNA fragment size standards for agarose gel electrophoresis were prepared by restriction endonuclease digestion of $\lambda\text{cI}857\text{Sam}7$ DNA (Boehringer Mannheim) with

EcoRI and *EcoRI/HindIII* or restriction endonuclease digestion of pPS359 (Section 4.2.3) with *HinfI*.

2.3.7 Isolation of DNA Fragments

Two methods were used for the recovery of specific DNA fragments from contaminating enzymes and other fragments of nucleic acid.

2.3.7.1 Electroelution from Agarose Gels

For isolation of a single restriction fragment from a restriction endonuclease digest and removal of enzymes from Bal31, Klenow end-filling and dephosphorylation reactions, a modification of the procedure of Dretzen *et al.*, (1981) was used. Fragments of DNA from an agarose gel were isolated by electroelution onto NA45 membrane (Schleicher and Schuell). The DNA was then recovered by elution into a solution of 1 M NaCl containing 50 mM arginine at 70 °C. After 1 h the membrane was removed and the DNA was precipitated with 2 volumes of ethanol at -70 °C. Following centrifugation (MSE Micro Centaur Microfuge [μfuge]), the pellet was dried briefly *in vacuo*, resuspended in TE and stored at -20 °C until required.

2.3.7.2 Extraction with Phenol/Chloroform

Phenol/chloroform extraction was utilised extensively in the preparation of single-stranded phage and single and double stranded phagemid DNA after treatment of phage particles with proteinase K and in removal of enzymes following Bal31, Klenow end-filling and dephosphorylation reactions. Extraction was carried out essentially as described (Sambrook *et al.*, 1989).

2.3.8 Ligation

Ligations were performed with T4 DNA ligase (Bresatech) using DNA fragments purified either by phenol/chloroform extraction (Section 2.3.7.2) or by electroelution from agarose gels (Section 2.3.7.1). Where ligations involved short double-stranded species (<100 bp), an insert to vector ratio of 10:1 was used to maximise cloning efficiency. Otherwise concentrations of the vector and inserts were adjusted on the basis of the theoretical considerations of Legerski and Robberson (1985) to optimise the production of the desired recombinant molecules.

2.3.8.1 Ligation of Blunt-ended Termini

Blunt-ended ligation reactions were carried out in the presence of $[\text{Co}(\text{NH}_3)_6]\text{Cl}_3$ and spermidine (Rusche and Howard-Flanders, 1985) using blunt-end ligation buffer (Sambrook *et al.*, 1989). To maximise intermolecular ligation, the cloning vector was dephosphorylated prior to ligation (Section 2.3.4). Reactions were routinely carried out at 30 °C for 2 h in volumes of 20 to 50 μl .

2.3.8.2 Ligation of Cohesive Termini and Recircularization Reactions

Ligations of cohesive termini created by restriction endonuclease digests were performed using the ligation buffer described by Sambrook *et al.* (1989). These ligations included intramolecular recircularizations, two- and three-fragment ligations and three-fragment ligations requiring only a single blunt-end ligation. Reactions were routinely carried out at 14 °C for 16 h in volumes of 20 to 50 μl .

2.3.9 Preparation of Competent Cells

Fresh competent *E. coli* were prepared using CaCl_2 by the procedure of Morrison (1979). All steps were carried out under sterile conditions at 0 °C. Cells were stored up to 6 months without substantial loss of competency in 15 % (v/v) glycerol at -70 °C.

2.3.10 Transformation of Competent *E. coli*

Transformation of *E. coli* was carried out with single- or double-stranded plasmid DNA as described (Morrison, 1979) except that heat treatment was routinely at 30 °C for 2 h. After cells had been allowed to express the antibiotic marker encoded by the plasmid DNA, they were spread as a lawn culture on an LBT plate containing the appropriate antibiotic. Colonies were usually visible after incubation at 30 °C for 16 h. Transformants were routinely purified by restreaking for single colonies prior to analysis.

2.3.11 3'-Radioactive End-Labeling of DNA

Radioactive labelling of DNA fragments enabled rapid screening and analysis of analytical preparations of recombinant plasmids prior to selection for DNA sequence determination. Appropriate restriction endonuclease digests were performed with ~0.1 µg of DNA to ensure target fragments contained suitable 5'-overhanging ends for incorporation of [α - 32 P]dATP. 3' end-labelling reactions were then carried out as described (Section 2.3.3) with replacement of dATP in the reaction mix with 4.4×10^7 Bq of [α - 32 P]dATP ($\sim 1.1 \times 10^{11}$ Bq/mmol, Amersham). Then 10 min after further addition of dATP (100 µM), the reaction was stopped by addition of one-half volume of RE Stop mix.

Separation and analysis of radioactively-labelled DNA fragments was carried out on 5 % polyacrylamide gels (30:2.7 acrylamide:bisacrylamide), polymerized with 0.033 % (w/v) ammonium persulfate and 0.021 % (v/v) TEMED in TBE buffer. Samples were loaded on slab gels (500 x 380 x 0.4 mm) mounted on a Bio-Rad Sequi-gen DNA sequencing apparatus containing TBE electrophoresis buffer. Electrophoresis was performed at 500 V using a Bio-Rad Model 3000 Xi electrophoresis power supply for the required time. Following electrophoresis, gels were affixed to Whatman 3MM filter paper, covered with plastic film and dried in a Bio-Rad Model 483 slab drier *in vacuo* at 80 °C over 1 h. Gels were then ready for autoradiography (Section 2.3.15). Sizes of products were determined by reference to the mobilities of standards (32 P-labelled *Sau3AI* fragments of pCE30; Elvin *et al.*, 1990).

2.3.12 Preparation of Single-Stranded DNA Templates

Single-stranded DNA used in site-directed mutagenesis, dideoxy sequencing and DNA replication assays was prepared using a modification of the procedure of Vieira and Messing (1987). All procedures were carried out at 4 °C unless otherwise indicated.

For preparation of M13 bacteriophage, a 1-l culture of *E. coli* strain K37 was grown at 37 °C in LBT to $A_{595} \sim 0.2$. This was infected with phage (10^9 pfu/ml) and growth was continued for 16 h at 37 °C with aeration. For preparation of phagemid particles, the *E. coli* strain AN2666 harbouring the phagemid was grown at 30 °C in LBT with 50 µg/ml ampicillin to an $A_{595} \sim 0.5$. The culture was then infected with M13K07 helper phage (5×10^9 pfu/ml) and growth continued for 16 h at 30 °C with aeration.

Phage particles were then prepared following the method of Hines and Ray (1980). Cultures were chilled, and cells were removed by centrifugation (7,600 x g, 30 min). At this point phage titres of the supernatant generally yielded ~ 2 to 3×10^{12} pfu/ml. Phage were precipitated over 2 h following the addition of 1/5 volume of 2.5 M NaCl, 25 % (w/v) polyethylene glycol-6000, and were harvested by centrifugation (7,600 x g, 30 min). The precipitate was passively resuspended in a volume 5 % that of the original culture with 0.5 % (v/v) N-lauryl sarcosinate (Sarkosyl) in TE and clarified by centrifugation (12,000 x g, 30 min). The phage were reprecipitated with polyethylene glycol as before; the pellet was passively resuspended in a minimum volume of TE. The suspension was clarified and the density of the supernatant adjusted by the addition of CsCl to 2.3 M. This was centrifuged in Beckman Ultra-Clear (14 x 89 mm) centrifuge tubes (100,000 x g, 48 h). The opalescent phage band was extracted from the tube using an 18-gauge hypodermic needle and syringe. The density of the phage solution was increased to 3.3 M by the addition of CsCl. It was overlaid in an Ultra-Clear centrifuge tube with 2.3 M CsCl in TE. The phage was isolated again by centrifugation as above and dialysed vs 2 changes of 1-l of TE over 12 h. Phage was routinely stored at 4 °C.

Single-stranded DNA was recovered from CsCl-density gradient purified phage following denaturation and digestion of the viral coat proteins. Purified phage were treated with 2 % (w/v) SDS at 60 °C for 5 min, then with 200 µg/ml Proteinase K (Boehringer Mannheim) at 37 °C for 30 min. Further Proteinase K was added to a final concentration of 300 µg/ml, and after a further 30 min at 37 °C, NaCl was added to 100 mM. The single-stranded DNA was isolated, following phenol/chloroform extraction, by precipitation with ethanol (Section 2.3.7.2).

The purity of the single-stranded DNA was assessed spectrophotometrically. A ratio of A_{260} to A_{280} near 1.8 indicated that the nucleic acid was essentially pure. DNA concentrations were determined spectrophotometrically assuming $A_{260} = 1$ is equivalent to a DNA concentration of 40 µg/ml. Single-stranded DNA was routinely stored at -70 °C.

2.3.13 Oligonucleotide Synthesis and Purification

Synthetic oligodeoxynucleotides used in dideoxy sequencing reactions and site-directed mutagenesis experiments were prepared by the A.N.U. Protein/DNA Facility using an Applied Biosystems 380B DNA Synthesizer. These were received in ammonia solution and were deprotected by treatment at 56 °C for 16 h prior to

concentration under vacuum at a rotary evaporator. After resuspension in mQH₂O, re-evaporation and resuspension in TE, oligonucleotides were stored at -20 °C until required.

2.3.14 Dideoxy Sequencing of Single- and Double-Stranded DNA

Determination of the nucleotide sequences of double-stranded plasmid DNA (Section 2.3.1.2) and single-stranded phage DNA (Section 2.3.12) was carried out using the Sanger dideoxy-mediated chain termination reaction (Sanger *et al.*, 1977). The procedure followed was essentially that of Tabor and Richardson (1987) using T7 DNA polymerase, with [α -³²P]dATP ($\sim 1.1 \times 10^{11}$ Bq/mmol, Amersham) for detection. Primers for sequencing reactions with pCE30 and derivatives are described (Figure 2.1). Other primers and their use are described in the text.

Analysis of sequence reactions was carried out on 6 % w/v polyacrylamide gels (19:1 acrylamide:bisacrylamide) in TBE buffer containing 8 M urea. Slab gels (500 x 380 x 0.4 mm) were mounted on a Bio-Rad Sequi-gen DNA sequencing apparatus containing TBE electrophoresis buffer and pre-electrophoresed for 30 min at 50-55 °C at 1500 V (Bio-Rad Model 3000Xi electrophoresis power supply). Following treatment at 80 °C for 2 min, samples were loaded and electrophoresis was continued for the required time. After electrophoresis, gels were affixed to 3MM filter paper, covered with plastic film and dried in a Bio-Rad Model 483 Slab Drier *in vacuo* at 80 °C over 1 h. Gels were then ready for autoradiography.

2.3.15 Autoradiography

Dried polyacrylamide gels containing [³²P]-radioactively labelled DNA were autoradiographed with XAR5 film (Kodak) in a light-proof cassette (Kodak) at ambient temperature for 6 to 12 hours. Intensifying screens were not used. Films were developed on a Kodak X-Omatic automatic X-ray film developer.

PROTOCOLS FOR PROTEIN MANIPULATION

2.4.1 FPLC Purification of Proteins

Protein purifications made use of a Fast Protein Liquid Chromatography (FPLC) System (Pharmacia). The major components of the system consisted of 2 High Precision P-500 pumps, a Liquid Chromatography Controller LCC-500 Plus, Automatic Motor Valves MV-7 and MV-8, a Fraction Collector FRAC-100, a single Path UV Monitor UV-M (280 nm filter) and a chart recorder (Rec 482). Blue dextran-Sepharose was prepared by CNBr activation of Sepharose 4B (Meyer *et al.*, 1979). Other ion exchange and gel filtration columns and their use are described in the text. Salt gradients were estimated by determining the conductivity of eluted samples with a PTI-58 Digital Conductivity Meter (Activon), with reference to standards prepared in the appropriate buffers.

2.4.2 Determination of Protein Concentration

Protein concentrations were determined by one of two methods. For quantification throughout protein purifications, for DNA replication assays and for densitometric analysis, the protein-dye binding method of Bradford (1976) was used with bovine serum albumin as standard. The concentration of proteins prepared for metal ion analysis were determined by measuring A_{280} . The extinction coefficient of the protein at 280 nm was estimated from the amino acid composition (Gill and von Hippel, 1989).

2.4.3 Denaturing SDS-Polyacrylamide Gel Electrophoresis

Denaturing SDS-polyacrylamide gel electrophoresis (SDS-PAGE) was used to ascertain cell protein profiles and protein purity, as described by Laemmli (1970). Slab gels for staining with silver (0.75 x 200 x 150 mm) or Coomassie blue (1.5 x 200 x 150 mm) were formed in vertical glass moulds. The resolving gels contained 9-12.5 % (w/v) acrylamide (as specified; 30:2.7 acrylamide:bisacrylamide), 375 mM Tris.HCl pH 8.8, 0.1 % (w/v) SDS, and were polymerised with 0.033 % (w/v) ammonium persulfate and 0.033 % (v/v) TEMED. Stacking gels contained 4.5 % acrylamide (30:2.7 acrylamide:bisacrylamide) 125 mM Tris.HCl pH 6.8, 0.1 % (w/v) SDS, and were polymerised with 0.08 % (w/v) ammonium persulfate and 0.08% (v/v) TEMED.

Electrophoresis tank buffer contained 51 mM Tris base, 384 mM glycine and 0.1 % (w/v) SDS. Protein samples were heated at 100 °C for 2 min in an equal volume of loading buffer (300 mM Tris base, 15 % (v/v) glycerol, 0.6 % (w/v) bromophenol blue, 50 mM DTT and 1 % (w/v) SDS) prior to loading. Gels were electrophoresed at 50 V (Bio-Rad 2000/02 power supply) until samples had migrated to the interface between the stacking and resolving gels. Electrophoresis was then continued at 120 V. M_r markers were phosphorylase b (93 kDa), bovine serum albumin (67 kDa), ovalbumin (43 kDa), carbonic anhydrase (30 kDa), soybean trypsin inhibitor (20.1 kDa), and α -lactalbumin (14.4 kDa) supplied in kit form (Pharmacia).

For staining with Coomassie blue, gels were fixed and stained in 40 % (v/v) methanol, 10 % (v/v) glacial acetic acid containing 0.3 % (w/v) Coomassie Blue R250 (Coomassie stain) for 2 to 6 h. The gels were destained in 10 % (v/v) propan-2-ol, 10 % (v/v) glacial acetic acid (destain solution). For staining with silver, gels were pretreated with gluteraldehyde, then stained with ammoniacal silver nitrate (Giulian *et al.*, 1983; See and Jackowski, 1989). Both Coomassie Blue and silver-stained gels were routinely photographed using a Model MP-4 Land camera (Polaroid) and Polaroid 4 x 5 type 55/high speed film. Films were developed according to the manufacturer's instructions. Photographs presented in this thesis were taken by Mr. B. Wight, RSC Photography Unit.

2.4.4 Densitometric Scanning

Densitometric analysis of Coomassie Blue and silver-stained SDS-PAGE gels made use of an LKB UltroScan XL Laser densitometer (Pharmacia). One-dimensional scans were performed with a 50 μ m x 2400 μ m rectangular Helium-Neon laser beam at 633 nm; the absorbance profile was recorded on an IBM personal computer graphics Epson 80 series dot matrix printer. The profiles from three separate scans were then used to determine peak areas; traces corresponding to peaks were cut from the paper charts and weighed.

2.4.5 Estimation of Protein Molecular Weight by Gel Filtration

Evaluation of the molecular weight of several proteins was done using gel filtration with a column (430 x 10 mm) of Sephacryl S-400 (Pharmacia) fitted with flow adaptors. Protein solutions were applied using a 200- μ l Sample Loop on the FPLC and eluted at 6 ml/h. The buffer used was 50 mM Tris.HCl pH 7.6, 20 % (v/v)

glycerol, 200 mM NaCl, 5 mM MgCl₂, 2 mM DTT, 0.1 mM ATP. Elution volumes (V_e) were determined from measurements of A_{280} . Protein standards were ribonuclease A (13.7 kDa), chymotrypsinogen A (25 kDa), ovalbumin (43 kDa), bovine serum albumin (67 kDa), aldolase (158 kDa), ferritin (440 kDa), and thyroglobulin (669 kDa) supplied in kit form (Pharmacia). K_{av} was determined as $(V_e - V_o)/(V_i - V_o)$, using V_o and V_i determined as V_e of blue dextran-2000 and ATP, respectively.

2.4.6 Metal Ion Analysis

The final compositions of buffers used for preparation of protein samples for metal ion analyses were 50 mM Tris.HCl pH 7.6, 20 % (v/v) glycerol, 5 mM EDTA and 1mM DTT (Buffer A); 50 mM Tris.HCl pH 7.6, 1 mM DTT (Buffer B). Protein samples were dialysed first against 1-l of Buffer A for 4 h then for 6 h vs 2 changes of 1-l of Buffer B. All dialyses were carried out at 4 °C. Glassware used in the preparation of samples was washed in chromic acid, boiled in 50 mM EDTA and washed and boiled in mQH₂O. Solutions were equilibrated at 25 °C before use.

All analyses were carried out using a Varian Spectra AA-30/40 automatic absorption spectrophotometer operated by Ms. R. Chao, ANU Microanalytical Unit.. Metal ion standards were made up in 4 % (v/v) nitric acid. For analyses of copper, iron, manganese and nickel, spectral lines of 240.7, 248.3, 279.5 and 232.0 nm, respectively, were used with a spectral band pass of 0.2 nm. For cobalt and zinc, spectral lines at 324.7 and 213.9 nm with a spectral band pass of 0.5 and 1.0 nm, respectively, were used. Samples were atomized in an air-acetylene flame. For iron, manganese and zinc analysis, hollow cathode lamps were operated at 5 mA, for cobalt and nickel at 3.5 mA and for copper at 7 mA. The final dialysis buffer alone had levels of metal ions below the limits of detection.

CHAPTER 3

Improved Overproduction and Purification of the *Escherichia coli* *dnaG* Primase, a Zinc Metalloprotein

The *Escherichia coli* primase is a 65.5-kDa DNA-dependent RNA polymerase. It is the enzyme believed responsible for synthesis of short RNA primers for DNA synthesis both on the leading strand at the chromosomal origin, *oriC* (van der Ende *et al.*, 1985) and, in association with the *dnaB* helicase and other components of the primosome (Arai *et al.*, 1981f), for repeated reinitiation on the lagging strand at replication forks (Kornberg and Baker, 1991). Primase is therefore an essential replication enzyme and its function and enzymology in the process of priming DNA for replication is now well understood. Studies of primer RNA synthesis on prokaryotic single-stranded phages have revealed four different modes of priming (Figure 1.3). Investigation of the interactions involved between primase and the phage G4 origin revealed structural features within the single-stranded template that appear to be essential for the function of a primase-dependent priming signal (Masai *et al.*, 1990a). It has been postulated that similar secondary structures may be engineered by the major replicative helicase, the *dnaB* protein, during replication directed by the ϕ X174-type primosome, the *oriC*-type primosome and in general priming (Arai *et al.*, 1981f). Based on its mobility and ability to multiply prime, the ϕ X174 type primosome was proposed to function on the lagging strand template for discontinuous DNA synthesis during *E. coli* chromosomal replication.

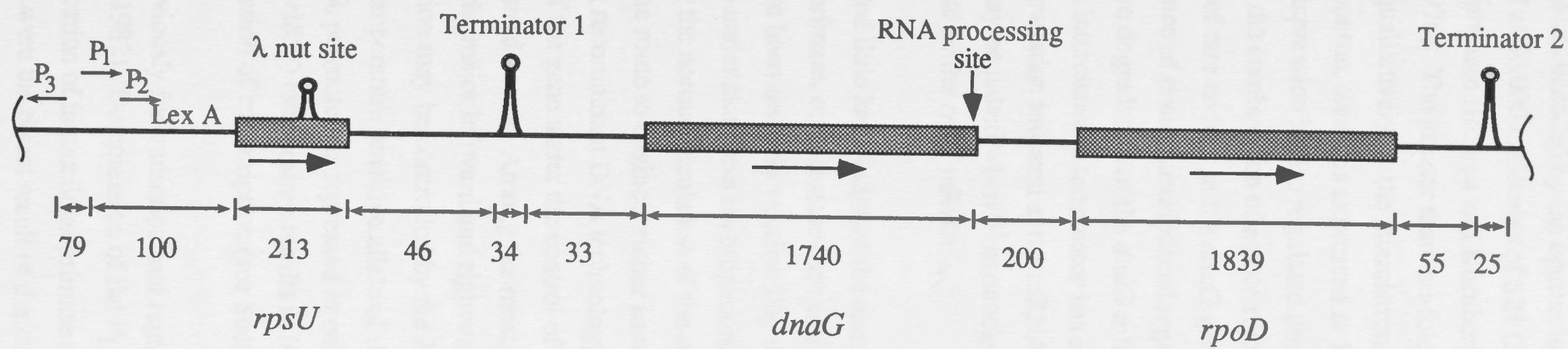
The *dnaG* gene, located at 66 minutes on the *E. coli* chromosome (Chen and Carl, 1975), is encoded within the macromolecular synthesis operon that also contains the *rpsU* and *rpoD* genes that encode ribosomal protein S21 and the sigma (σ) subunit of RNA polymerase holoenzyme, respectively (Lupski and Godson, 1984; Figure 3.1). This operon may allow *E. coli* cells the opportunity to coordinate regulation of replication, transcription and translation through control of expression of genes required at the initiation of each of these processes. The operon itself contains two terminators, an internal terminator in the intergenic region between *rpsU* and *dnaG* and a terminator which ends the operon (Burton *et al.*, 1983; Lupski *et al.*, 1983). Expression of the *dnaG* and *rpoD* genes requires suppression of the terminator upstream of the ATG start codon of the *dnaG* gene. This transcriptional attenuation has been demonstrated in the presence of the λ N gene product (pN) which binds to a λ nut site-equivalent within the *rpsU* gene (Nakayama and Yura, 1976; Lupski *et al.*, 1983).

It has been postulated that the existence of an *E. coli* antiterminator protein, analagous in function to the λ N gene product pN, would therefore allow transcription initiated from a promoter upstream of the *rpsU* gene to proceed through the terminator and synthesize both *dnaG* and *rpoD* mRNA (Smiley *et al.*, 1982). Such a system would

Figure 3.1

Arrangement of the regulatory units of the *rpsU-dnaG-rpoD* macromolecular synthesis operon. Arrows indicate the direction of transcription and the numbers below the line represent in base pairs the sizes of the genes and regions between the genes and regulatory units. It is proposed that suppression of the RNA polymerase *rho*-independent transcriptional termination signal upstream of the ATG translation initiation codon of *dnaG* is executed by an *E. coli* antiterminator protein, analogous in function to the λN gene product pN. The binding of such a protein to a *nut*-like sequence in the *rpsU* gene may allow transcription initiated from the promoter further upstream to proceed through the terminator and synthesize *dnaG-rpoD* mRNA. The *dnaG* gene is further regulated by mRNA processing commencing downstream of the *dnaG* sequence and probably halted by the presence of the intergenic terminator (Terminator 1). P₁ and P₂ represent the promoters identified as responsible for directing transcription of the operon. A strong *lexA* binding site implicates the *rpsU-dnaG-rpoD* macromolecular synthesis operon in the SOS regulatory system.

Figure adapted from Lupski *et al* (1983).



be analagous to control of early gene function in bacteriophage λ (Ward and Gottesman, 1982), but no such protein has yet been identified.

Steady state levels of the three proteins encoded by the operon in exponentially growing cells have been estimated at 50,000 molecules of S21 (Kjeldgaard and Gausing, 1974), 50 molecules of primase (Rowen and Kornberg, 1978) and 3,000 molecules of σ (Engbaek *et al.*, 1976). The greater than ten-fold higher expression of *rpsU* than *rpoD* can be explained qualitatively by the intercistronic terminator.

However, the middle gene in the operon, *dnaG*, is expressed at 1/60 the level of the distal gene, *rpoD*. Clearly *dnaG* expression is also regulated post-transcriptionally. Smiley *et al* (1982) suggested that the combination of a poor ribosome-binding site (Figure 3.3) and a high frequency of rare codons in the *dnaG* gene decrease *dnaG* mRNA translation. There is evidence of post-transcriptional regulation of primase levels through a process of selective degradation of the *dnaG* mRNA (Burton *et al.*, 1983). This might suggest that the intercistronic terminator has a second role in gene regulation in this operon. Any degradation initiated at the mRNA processing site located between *dnaG* and *rpoD* may be halted when 3' exonucleases encounter the stem and loop structure at the 3' end of the *rpsU* mRNA.

Enzymological studies have provided little knowledge of the chemistry involved in primer formation, the structure of primase, or its interactions with DNA and other proteins. Studies such as these have been severely restricted by the scarcity of primase in normal *E. coli* cells. To undertake these investigations it was necessary to overcome the processes controlling the normal regulation of the *dnaG* gene to provide enriched sources of the protein. One route to engineer richer sources is overproduction of the protein using recombinant DNA technology. This can be achieved by placing transcription of the gene under the control of a strong inducible *E. coli* promoter in an appropriate plasmid vector. Among the most useful vectors for this purpose are those that contain the major leftward and rightward promoters (P_R or P_L) of bacteriophage λ . Transcription may be controlled by the λ repressor, supplied by expression of the *cI* gene. If a temperature sensitive allele of the *cI* gene is used (e.g. *cI857*) transcription from the λ promoters is repressed in cells growing at 30 °C. Placing cultures at a restrictive (for *cI857*) temperature results in inactivation of the thermolabile repressor and derepression of transcription (see Section 2.1.2).

Such a system has been utilized previously for transcriptional regulation of the cloned *dnaG* gene (Wold and McMacken, 1982). Derepression of the P_L promoter on these plasmids elicited little or no amplification of intracellular primase concentrations. It was postulated that these low levels were the direct result of the *rho*-independent transcriptional terminator upstream of the *dnaG* start codon. However, amplification

of primase levels was achieved by increasing the gene dosage of *dnaG* (Wold and McMacken, 1982). Using a bacteriophage λ -pBR322 chimaeric plasmid, whose copy number could be increased by thermal induction (Rao and Rogers, 1978), primase levels were amplified 100-fold over levels in wild-type *E. coli*. This has made purification of primase at levels adequate for enzymological studies feasible.

It is evident from these results that considerable improvement in levels of overproduction of primase would require the replacement of DNA sequences responsible for the normal control of its expression. Dr N.E. Dixon and P.E. Lilley achieved this by using exonuclease Bal31 to delete DNA upstream of the *dnaG* ATG translation initiation codon that includes the transcriptional terminator. The gene was then manipulated further to maximize transcription of the gene and increase the efficiency of translation from the message, by providing it with a strong ribosome-binding site and placing it under control of the tandem strong bacteriophage λ promoters in the vector pCE30 (Elvin *et al.*, 1990). The successful overproduction of soluble primase from this source has enabled the purification of primase in quantities adequate for studies of its structure, its interactions with other replication proteins and the DNA template, and the mechanism and chemistry of primer production.

MATERIALS AND METHODS

3.2.1 Amplification of *dnaG* Primase by Gene Dosage

Plasmid pPL184 is a pCE30 derivative (see Figures 3.2, 3.3 and 3.6).¹ Plasmid pPS156 was constructed by insertion of the 1999-bp *Bam*H I fragment from pPL184 bearing *dnaG* and the synthetic ribosome-binding site into the λ promoter-runaway replication vector pCE33 (Elvin *et al.*, 1990). Transformants of AN1459 were selected on plates containing chloramphenicol. Plasmids of the anticipated size were further screened for correct orientation of the *dnaG* gene by *Bam*HI and *Hind*III/*Mlu*I digestion of DNA from small-scale analytical plasmid preparations. The plasmid

1. P.E. Lilley and N.E. Dixon, unpublished.

construct pPS156 was selected from a group of plasmids that directed the expression of the *dnaG* gene product at 42 °C.

3.2.2 The Primase Dependent G4 SS→RF Replication Assay

Sources of *E. coli* replication proteins were: Highly-purified single-stranded DNA binding protein (8.8×10^4 units/mg)² and the β subunit of DNA polymerase III holoenzyme (3×10^6 units/mg)³ were prepared from new sources and by modified procedures, and were gifts from Drs P. Hendry and J.L. Beck (Research School of Chemistry, Australian National University). DNA polymerase III* (5.5×10^5 units/mg), purified through Fraction IV (Maki *et al.*, 1988), was a gift from Dr N.E. Dixon (Research School of Chemistry, Australian National University). The *dnaG* primase (1.36×10^6 units/mg) was prepared as described in this chapter. R199/*G4ori_c* (Sakai and Godson, 1985) was a gift from Dr A. Kornberg (Stanford University).

The G4 SS→RF replication assay, with R199/*G4ori_c* single-stranded DNA as template, was essentially as described (Maki *et al.*, 1988). Reaction mixtures (25 μ l) contained 20 mM Tris.HCl, pH 8.0, glycerol (4 % v/v), dithiothreitol (8 mM), bovine serum albumin (80 μ g/ml), magnesium acetate (8 mM), ATP (2 mM), CTP, GTP and UTP (100 μ M each), dATP, dCTP and dGTP (50 μ M each), [³H]-TTP (50 μ M, 3.3 Bq/pmol), single-stranded DNA template (230 pmol, as nucleotide), single-stranded DNA binding protein (600 ng), DNA polymerase III* (100 units) and the β subunit (16 ng). The mixtures were assembled at 0 °C then, after addition of primase, were treated for 5 min at 30 °C. DNA synthesis was terminated by cooling at 0 °C and addition of trichloroacetic acid (10 % w/v, containing 0.1 M NaPP_i) Acid precipitates were collected on Whatman GF/C filter discs, which were washed with 1 N HCl containing 0.1 M NaPP_i and ethanol, dried and counted in a liquid scintillation counter. One unit of enzymatic activity is the amount that promotes incorporation of one pmol of nucleotide into product in one minute.

2. P. Hendry and N.E. Dixon, unpublished.

3. J.L. Beck and N.E. Dixon, unpublished.

3.2.3 Purification of Primase

The final composition of the lysis buffer was: 50 mM Tris.HCl pH 7.5, 10 % (w/v) sucrose, 0.15 M NaCl, 1 mM dithiothreitol, 20 mM EDTA, 20 mM spermidine.3HCl, 0.1 % (w/v) Brij-58. Buffers used for chromatography were 10 mM MES.NaOH pH 6.75 (Buffer A) and 20 mM Tris.HCl pH 7.6 (Buffer B). Both buffers also contained glycerol (20 % v/v), EDTA (1 mM) and dithiothreitol (2 mM). All procedures during purification of primase were at 0 - 5 °C.

Strain RSC310 (AN1459 containing plasmid pPL195; Figure 3.2), was grown at 30 °C in LB broth (38 l) containing 50 mg/l ampicillin in a fermentor (New Brunswick Scientific) to $A_{595} = 0.65$. After a rapid temperature shift to 42 °C, growth was continued at 42 °C for a further period of 4 h (final $A_{595} = 2.0$). The culture was chilled; cells (106.5 g, wet weight) were harvested (Sharples centrifuge), resuspended in 320 ml of 67 mM Tris.HCl pH 8.1, 13.3 % (w/v) sucrose, poured into liquid N₂, and stored at -70 °C.

RSC310 cell suspension (57 g, from 5 l of cell culture) was thawed and diluted to 425 ml to the final composition of lysis buffer. Lysozyme (0.2 mg/ml) was added, and the mixture was stirred for 1 h at 0 °C. To the supernatant (Fraction I, 420 ml) obtained after centrifugation (40,000 x g, 30 min) was added solid ammonium sulfate (0.24 g/ml). After being stirred for 1 h, the suspension was centrifuged (40,000 x g, 30 min). The pellet was dissolved in 30 ml of Buffer A, and the solution was dialysed for 2 h vs. two changes (800 ml each) of Buffer A (Fraction II). Fraction II was diluted to 154 ml with Buffer A, and applied to a column (2.5 x 12 cm) of Bio-Rex 70 cation exchange resin pre-equilibrated in Buffer A. The column was washed with Buffer A (120 ml); bound protein was eluted with a linear gradient (500 ml) of 0 to 3 M NaCl in Buffer A at a flow rate of 45 ml/h (Figure 3.8). Fractions containing primase were pooled (Fraction III, 62 ml). Fraction III was dialysed over 3 h vs. two changes (1 l) of Buffer B, diluted to 200 ml with Buffer B, and applied to a column (2.5 x 16 cm) of DEAE-Sephacel anion-exchange resin pre-equilibrated in Buffer B. The column was washed with Buffer B (160 ml); bound proteins were eluted with a gradient (800 ml) of 0 to 1 M NaCl in Buffer B at a flow rate of 45 ml/h (Figure 3.9). Fractions containing primase were pooled (Fraction IV, 45 ml).

RESULTS

Overproduction of dnaG primase

The strategy used to obtain overproduction of primase beginning with the *dnaG* gene on a DNA fragment from pRLM61 is summarized in Figure 3.2. Because the extent of digestion with Bal31 could not be controlled precisely, a family of plasmids was produced that differed in the distance between the ribosome-binding site and ATG translation initiation codon. Strains containing these plasmids were selected and screened individually to estimate this distance, and four different plasmid constructs were identified. Three strains containing plasmids that differed in spacing between the ribosome-binding site and the normal *dnaG* translation initiation codon overproduced primase to similarly high levels (Figure 3.3). Within the narrow limits examined here it appears that variation of this spacing has no greater than a two-fold effect on *dnaG* expression. The fourth strain (RSC291) overproduced a slightly smaller protein consistent with the results of nucleotide sequencing that indicated that the new ribosome-binding site was fused sufficiently close to the GTG codon that normally encodes Val-27 of primase that it ought to become a new translation initiation codon. Although soluble extracts from RSC291 contained appreciable amounts of the N-terminally detected protein (primase- $\Delta 27$, M_r 62,500), they had no detectable primase activity (Figure 3.4). These data suggest that the NH₂-terminal portion of primase has an important role in determination of its activity.

Primase is essentially stable at 42 °C (Bouché *et al.*, 1975) and inactivation of the protein during thermal induction was not anticipated. A major concern with primase overproduced from these sources, however, was that overexpression of a gene that contains many rare codons might result in misincorporation of non-cognate amino acids into the protein, thereby compromising its enzymic activity. This was investigated in a limited way by overproducing the *dnaY* (*argU*) tRNA (Garcia *et al.*, 1986) in the same cells with primase (Figure 3.5). High-level expression of many recombinant genes in *E. coli* have been shown to be dependent on the availability of the *dnaY* gene product (Brinkmann *et al.*, 1989), however, neither the level of primase overproduction nor the specific activity of the purified proteins was improved in its presence (Figure 3.5).

An attempt was made further to improve overproduction of primase by increasing gene dosage in the cell. This was achieved by inserting a fragment of pPL184 bearing *dnaG* and the synthetic ribosome-binding site into the λ promoter-runaway replication

Figure 3.2

Construction of a plasmid, pPL195, that directs overproduction of primase. A 3434-bp *Hind*III/*Bam*HI fragment from pRLM61 containing all of *dnaG* and 180 bp of DNA upstream of the ATG translation initiation codon was isolated and treated with sufficient exonuclease Bal31 to remove ~175 bp from each end. After repair of the ends by treatment with the large fragment of DNA polymerase I, product fragments were digested with *Pvu*II and ligated with an ~100-fold molar excess of 5'-phosphorylated self-complementary 32-bp *Bam*HI-ribosome-binding site linker (5'-TTAAACCTCCTTAGGATCCTAAGGAGGTTTAA^{-OH}). Following digestion with *Bam*HI the *dnaG*⁺ fragment (2.0 kb) was isolated and ligated to 5'-dephosphorylated *Bam*HI-linearized pCE30 (Figure 2.1). Ampicillin resistant transformants were selected at 30 °C. On the assumption that high-level overexpression of *dnaG* would affect their viability, strains were first screened for temperature sensitivity on plates at 42 °C. From selected temperature-sensitive transformants, plasmids of the anticipated size (~6.0 kb) were further screened by *Bam*HI/*Mlu*I restriction endonuclease digests and 3'-end labelling studies of *Bam*HI/*Mlu*I fragments from small-scale analytical plasmid preparations to determine the extent of exonuclease digestion by Bal31. Plasmids predicted to possess short segments of DNA between the ATG translation initiation codon of the *dnaG* gene and the synthetic ribosome-binding site of the attached 32-bp linker were sequenced in the region of the 5' end of *dnaG* using the dideoxy method. The relevant sequences of these plasmids are given in Figure 3.4. Selected strains were then analysed for expression of the desired gene product at 42 °C by visualization using SDS-PAGE (Figure 3.4). The *Mlu*I sites used in screening of plasmids are centred 20 and 112 bp from the 5' end of the *dnaG* coding region.

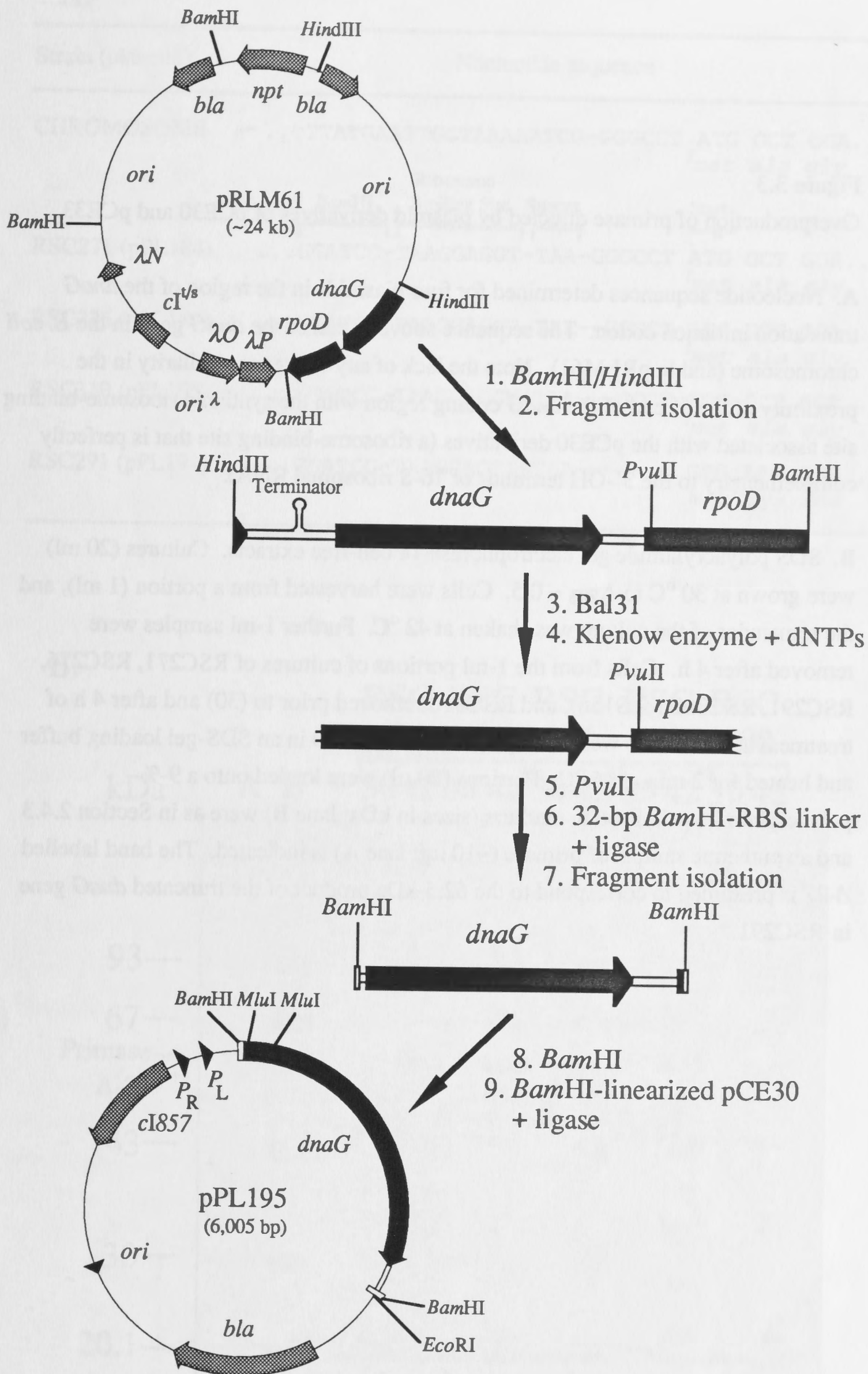


Figure 3.3

Overproduction of primase directed by plasmid derivatives of pCE30 and pCE33.

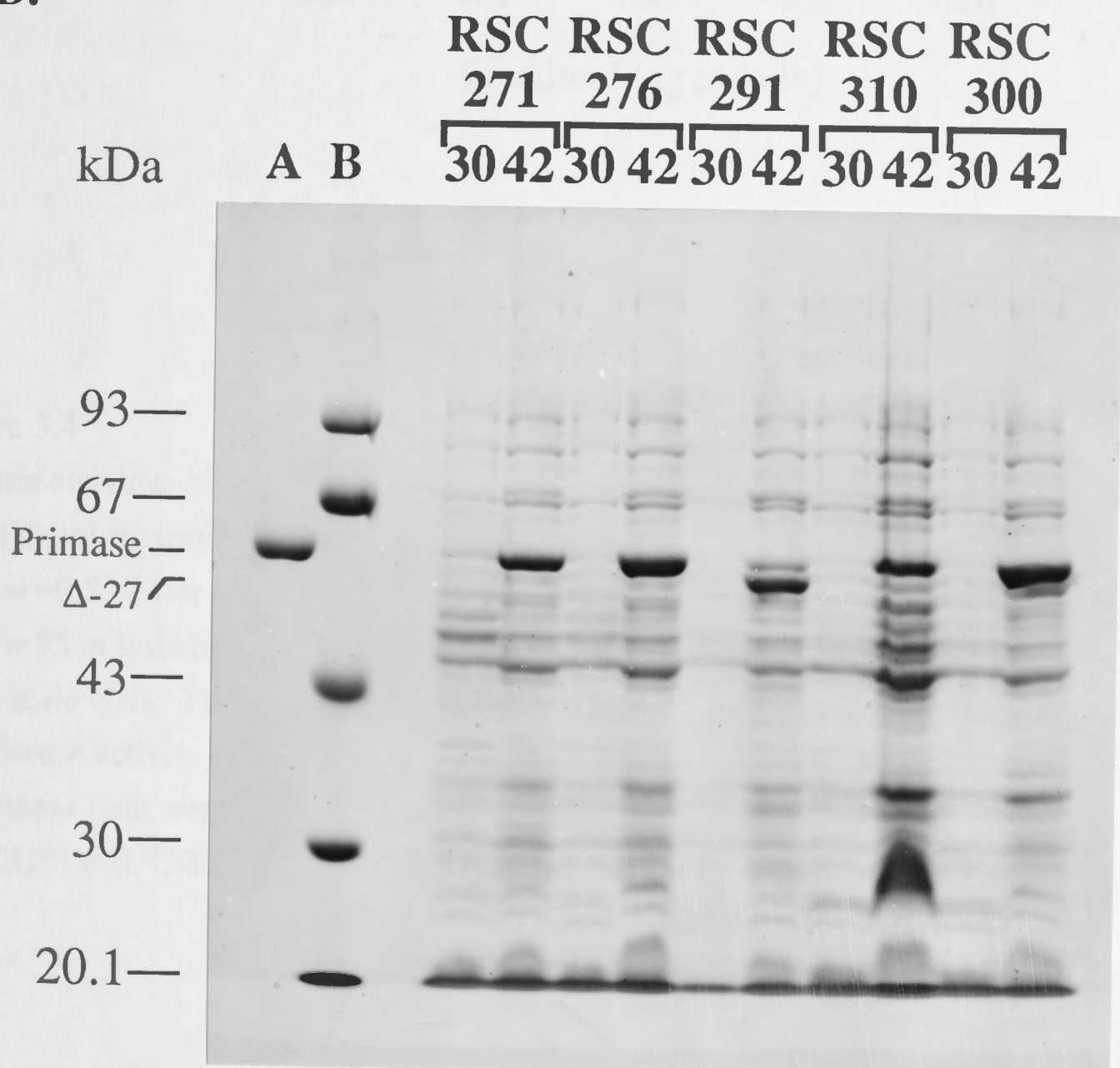
A. Nucleotide sequences determined for four plasmids in the region of the *dnaG* translation initiation codon. The sequence above is that of the *dnaG* gene in the *E. coli* chromosome (and in pRLM61). Note the lack of any sequence similarity in the proximity of the 5' end of the *dnaG* coding region with the synthetic ribosome-binding site associated with the pCE30 derivatives (a ribosome-binding site that is perfectly complementary to the 3'-OH terminus of 16-S ribosomal RNA).

B. SDS polyacrylamide gel electrophoresis of cell-free extracts. Cultures (20 ml) were grown at 30 °C to $A_{595} \sim 0.5$. Cells were harvested from a portion (1 ml), and the remainder of the culture was shaken at 42 °C. Further 1-ml samples were removed after 4 h. Cells from the 1-ml portions of cultures of RSC271, RSC276, RSC291, RSC300 (pPS156), and RSC310 removed prior to (30) and after 4 h of treatment at 42 °C (42) were resuspended to $A_{595} = 10$ in an SDS-gel loading buffer and heated for 2 min at 95 °C. Portions (20 μ l) were loaded onto a 9-% polyacrylamide gel. Protein markers (sizes in kDa; lane B) were as in Section 2.4.3 and an authentic sample of primase ($\sim 10 \mu$ g; lane A) is indicated. The band labelled Δ -27 is presumed to correspond to the 62.5-kDa product of the truncated *dnaG* gene in RSC291.

A.

Strain (plasmid)	Nucleotide sequence	
CHROMOSOME	...TTATGAATTGCTAAAAATCG-GGGCCT	ATG GCT GGA... ^l met ala gly
	<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;"> <i>Bam</i>HI ┌───┐ └───┘ </div> <div style="text-align: center;"> Ribosome- Binding Site ┌───┐ └───┘ </div> <div style="text-align: center;"> Spacer ┌───┐ └───┘ </div> <div style="text-align: center;"> Start → </div> </div>	
RSC271 (pPL184)	...GGATCC-TAAGGAGGT-TAA-GGGCCT	ATG GCT GGA... ^l met ala gly
RSC276 (pPL189)	...GGATCC-TAAGGAGGT-TAA--GGCCT	ATG GCT GGA... ^l met ala gly
RSC310 (pPL195)	...GGATCC-TAAGGAGGT-TAA---GCCT	ATG GCT GGA... ^l met ala gly
RSC291 (pPL191)	...GGATCC-TAAGGAGGT-TAA-----	GTG AAG CTG... met ²⁸ lys leu

B.



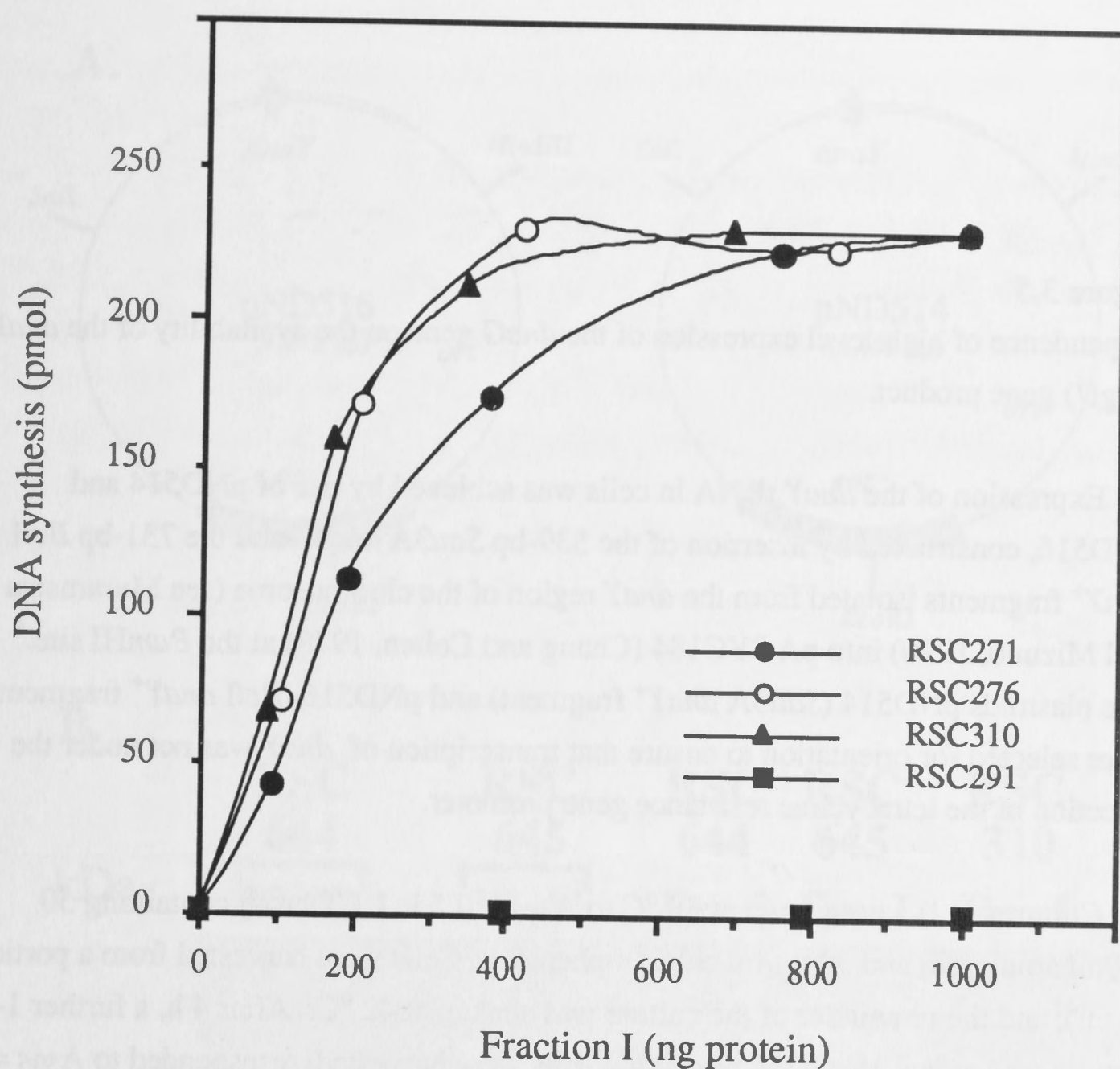


Figure 3.4

Primase activities of extracts from thermally induced strains containing plasmids that direct the overexpression of the *dnaG* gene. Cultures (20 ml) were grown at 30 °C to $A_{595} = \sim 0.5$. After 2 h, cells were harvested from 10 ml of culture, resuspended to $A_{595} = 25$ in lysis buffer and treated essentially as described in Section 3.2.3 for large-scale lysis. The supernatant following centrifugation (Fraction I) was assayed for primase activity (DNA synthesis) and for protein. Specific activities calculated from these data were 1.19×10^5 (RSC271), 1.65×10^5 (RSC276), 1.83×10^5 (RSC310) and <200 (RSC291) units/mg.

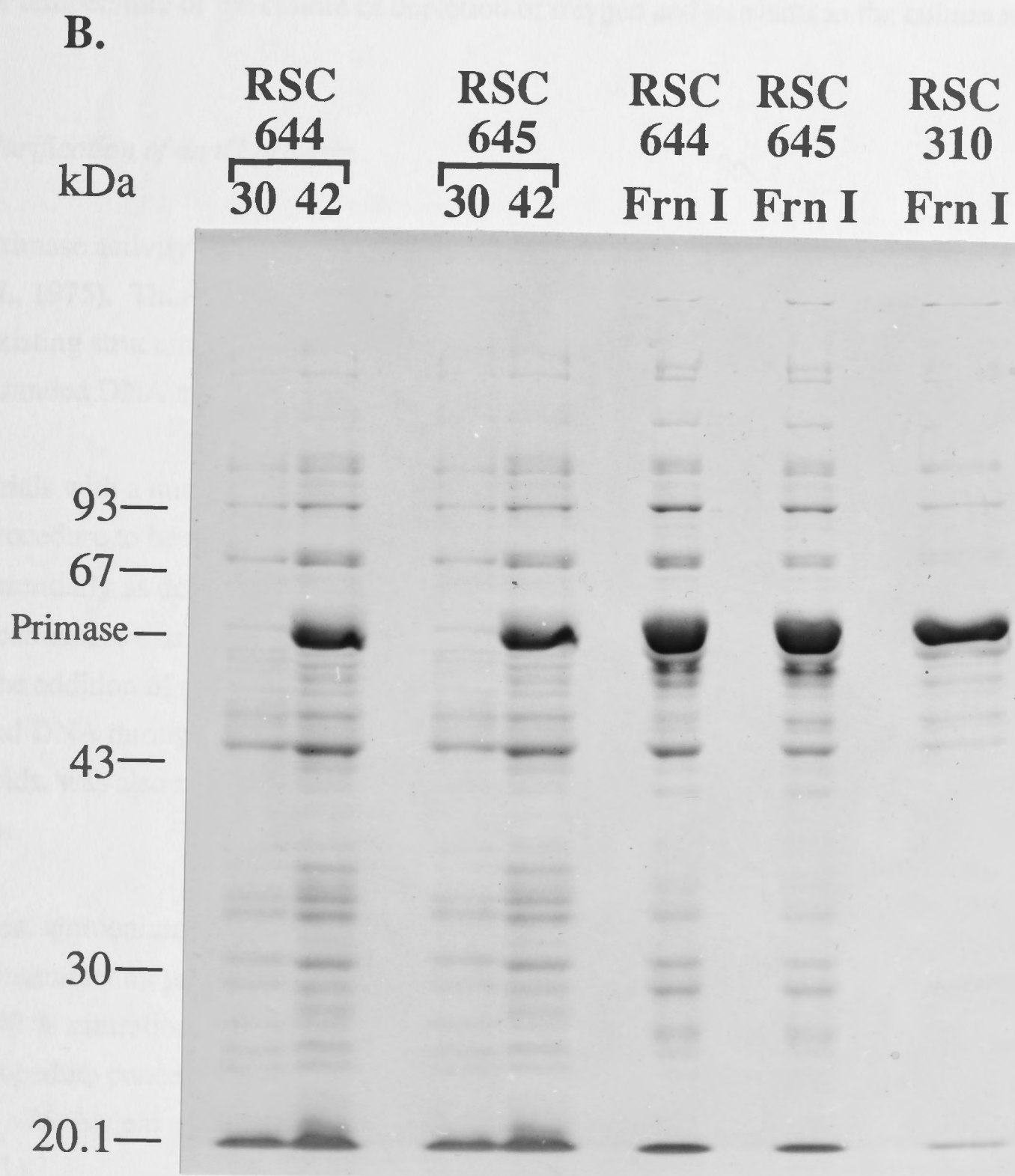
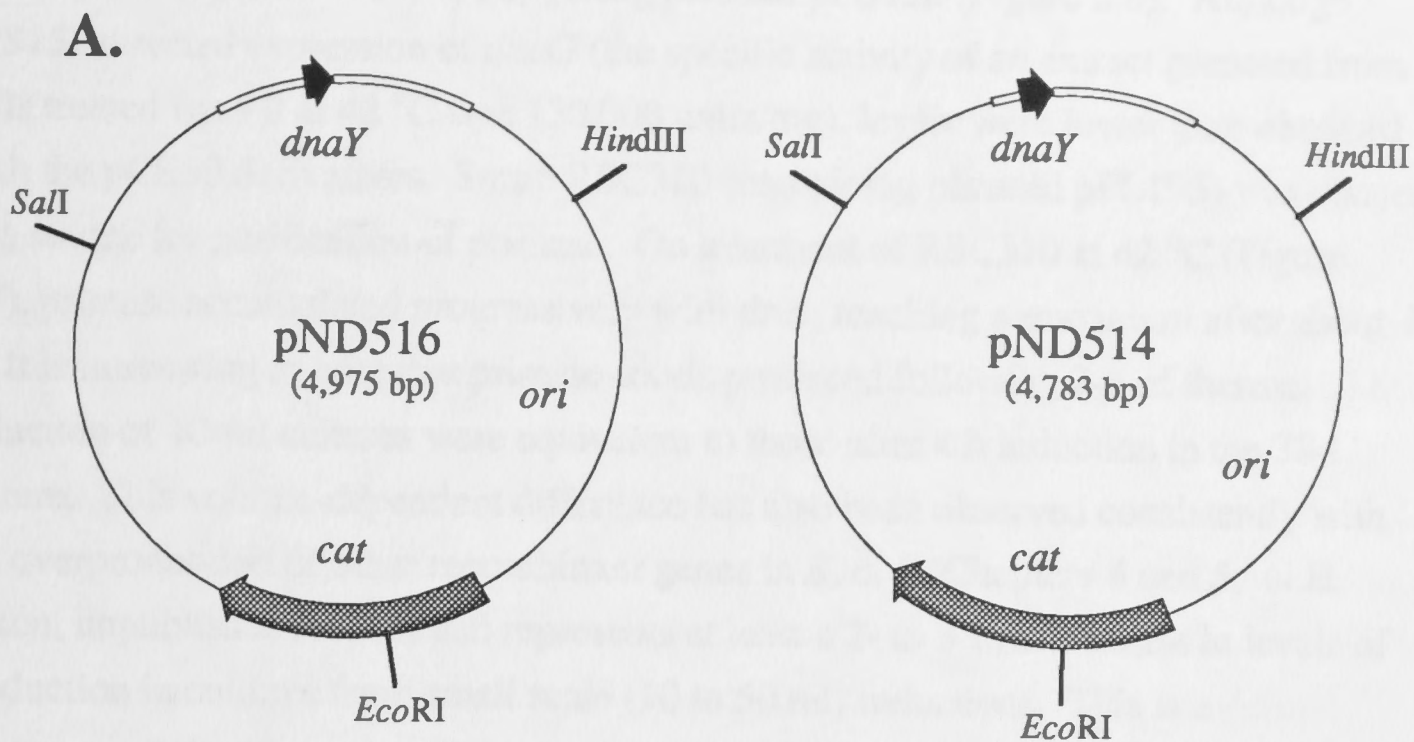
Figure 3.5

Dependence of high-level expression of the *dnaG* gene on the availability of the *dnaY* (*argU*) gene product.

A. Expression of the *dnaY* tRNA in cells was achieved by use of pND514 and pND516, constructed by insertion of the 539-bp *Sau3A dnaY*⁺ and the 731-bp *BclI dnaY*⁺ fragments isolated from the *dnaY* region of the chromosome (see Muramatsu and Mizuno, 1990) into pACYC184 (Chang and Cohen, 1978) at the *Bam*HI site.⁴ The plasmids pND514 (*Sau3A dnaY*⁺ fragment) and pND516 (*BclI dnaY*⁺ fragment) were selected for orientation to ensure that transcription of *dnaY* was not under the direction of the tetracycline resistance gene promoter.

B. Cultures (1 l) were grown at 30 °C to A₅₉₅ ~ 0.5 in LBT broth containing 50 µg/ml ampicillin and 50 µg/ml chloramphenicol. Cells were harvested from a portion (1 ml), and the remainder of the culture was shaken at 42 °C. After 4 h, a further 1-ml sample was removed and the remaining cells were harvested, resuspended to A₅₉₅ = 25 in lysis buffer and treated essentially as described (Section 3.2.3) for large-scale lysis. The supernatant following centrifugation (Fraction I) was assayed for primase activity and for protein. Specific enzymatic activities calculated from these data were 1.41 × 10⁵ (RSC644; pND514 + pPL195), 1.71 × 10⁵ (RSC645; pND516 + pPL195) and 1.47 × 10⁵ (RSC310; pPL195) units/mg. Cells from the 1-ml portions of cultures of RSC644 and RSC645 removed prior to (30) and after 4 h of treatment at 42 °C (42) were resuspended to A₅₉₅ = 10 in an SDS-gel loading buffer and heated for 2 min at 95 °C. Portions (20 µl) were loaded onto a 9-% polyacrylamide gel along with similarly treated samples (containing 10,000 units of primase) of Fractions I (Frn I) of all three preparations. Protein markers (sizes in kDa) were as in Section 2.4.3. Densitometric scanning of a Coomassie Blue R250 stained gel (Section 2.4.4) loaded with these samples indicated that each band corresponding to 10,000 units of primase contained identical amounts of the enzyme.

4. N.E. Dixon, unpublished.



vector pCE33 (Elvin *et al.*, 1990) giving plasmid pPS156 (Figure 3.6). Although pPS156 directed expression of *dnaG* (the specific activity of an extract prepared from cells treated for 4 h at 42 °C was 130,000 units/mg), levels were lower than obtained with the pCE30 derivatives. Strain RSC310 (containing plasmid pPL195) was chosen as a source for purification of primase. On treatment of RSC310 at 42 °C (Figure 3.7), primase accumulated progressively with time, reaching a maximum after about 4 h. It is interesting to note that primase levels produced following 2-h of thermal induction of 10-ml cultures were equivalent to those after 4 h induction in the 38-1 culture. This volume-dependent difference has also been observed consistently with the overproduction of other recombinant genes in *E. coli* (Chapters 4 and 5; N.E. Dixon, unpublished results) and represents at least a 2- to 3-fold increase in levels of production in cultures from small scale (10 to 50 ml) inductions. This is a phenomenon that does not seem to be a direct result of small differences in either pH or temperature of the culture or depletion of oxygen and nutrients in the culture media.

Purification of dnaG primase

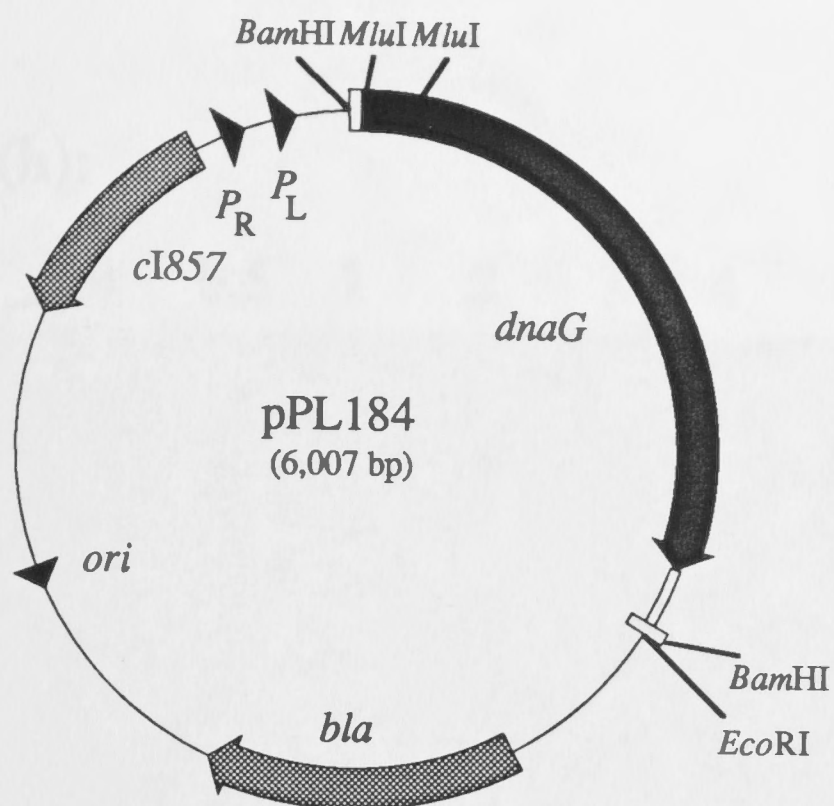
Primase activity was assessed with the primase-specific G4 SS→RF assay (Bouché *et al.*, 1975). This primer extension assay is dependent upon primase recognizing a pre-existing structure at the complementary strand origin of SSB-coated G4 single-stranded DNA at which primase produces a unique primer (Figure 1.3).

Trials with a number of different lysis techniques showed the most efficient lysis procedure to be the use of lysozyme in the presence of the nonionic detergent, Brij-58, essentially as described (Rowen and Kornberg, 1978). This procedure enabled almost quantitative extraction of overproduced primase into a soluble fraction (Fraction I). The addition of spermidine to the lysis reaction facilitated the removal of cell debris and DNA during centrifugation. This reagent which binds to and compacts nucleic acids, was also used in all other preparations of replication enzymes (Chapters 5 and 6).

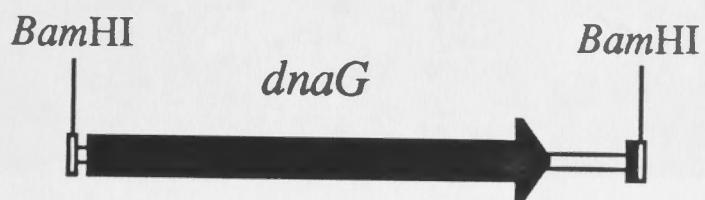
Trial ammonium sulfate precipitation of primase in Fraction I indicated that the bulk of contaminating protein could be removed with addition of solid ammonium sulfate (to ~40 % saturation) essentially as described (Rowen and Kornberg, 1978). This procedure concentrated the soluble primase solution prior to column chromatography in >75 % yield and was accompanied by a 2.6-fold enrichment of the protein (Table 3.1).

Figure 3.6

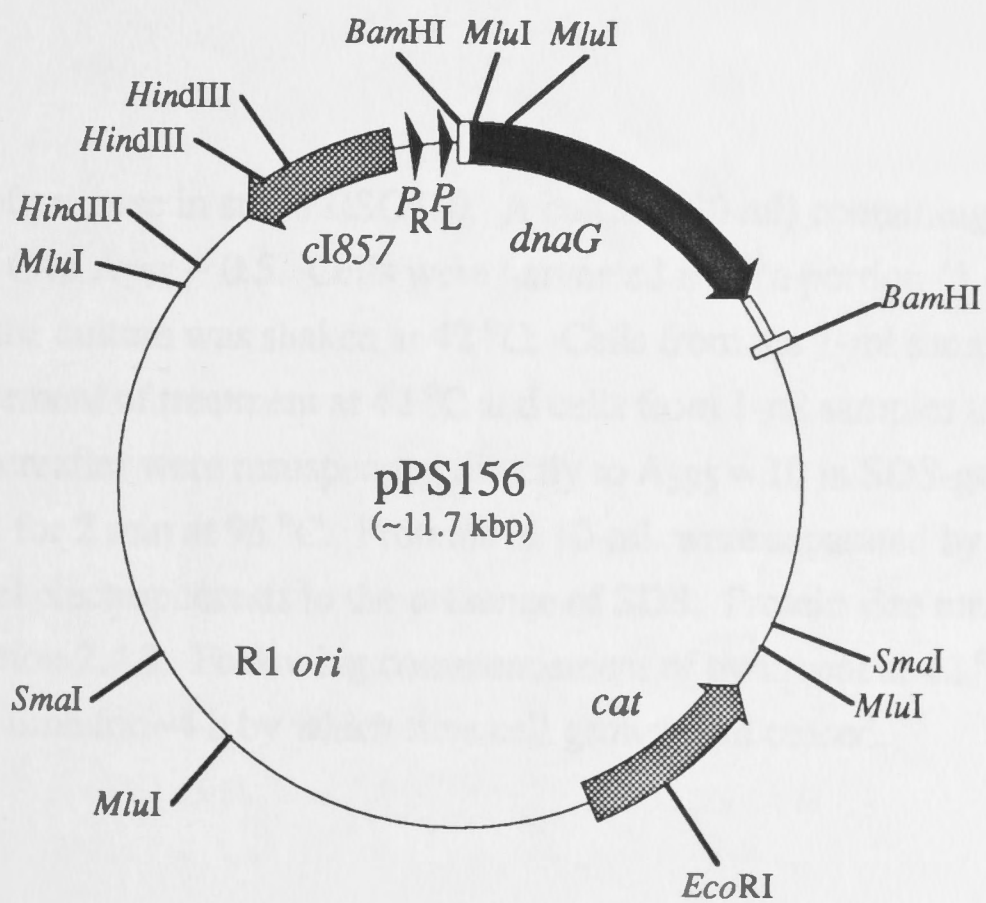
Construction of a plasmid, pPS156, that directs overexpression of the *dnaG* gene coupled with gene amplification. Plasmid pPS156 was constructed by insertion of the 2.0-kb *Bam*HI fragment of pPL184 (Figure 3.3) bearing *dnaG* and the synthetic ribosome-binding site into the *Bam*HI site of the vector pCE33 (Section 2.1.2). The *Mlu*I sites used in screening of plasmids are centred 20 and 112 bp from the 5' end of the *dnaG* coding region. The *Hind*III site of the vector that was similarly used is located 791 bp upstream of the ATG translation initiation codon of the *dnaG* gene.



1. *Bam*HI
2. Fragment isolation



3. *Bam*HI-linearized pCE33 + ligase



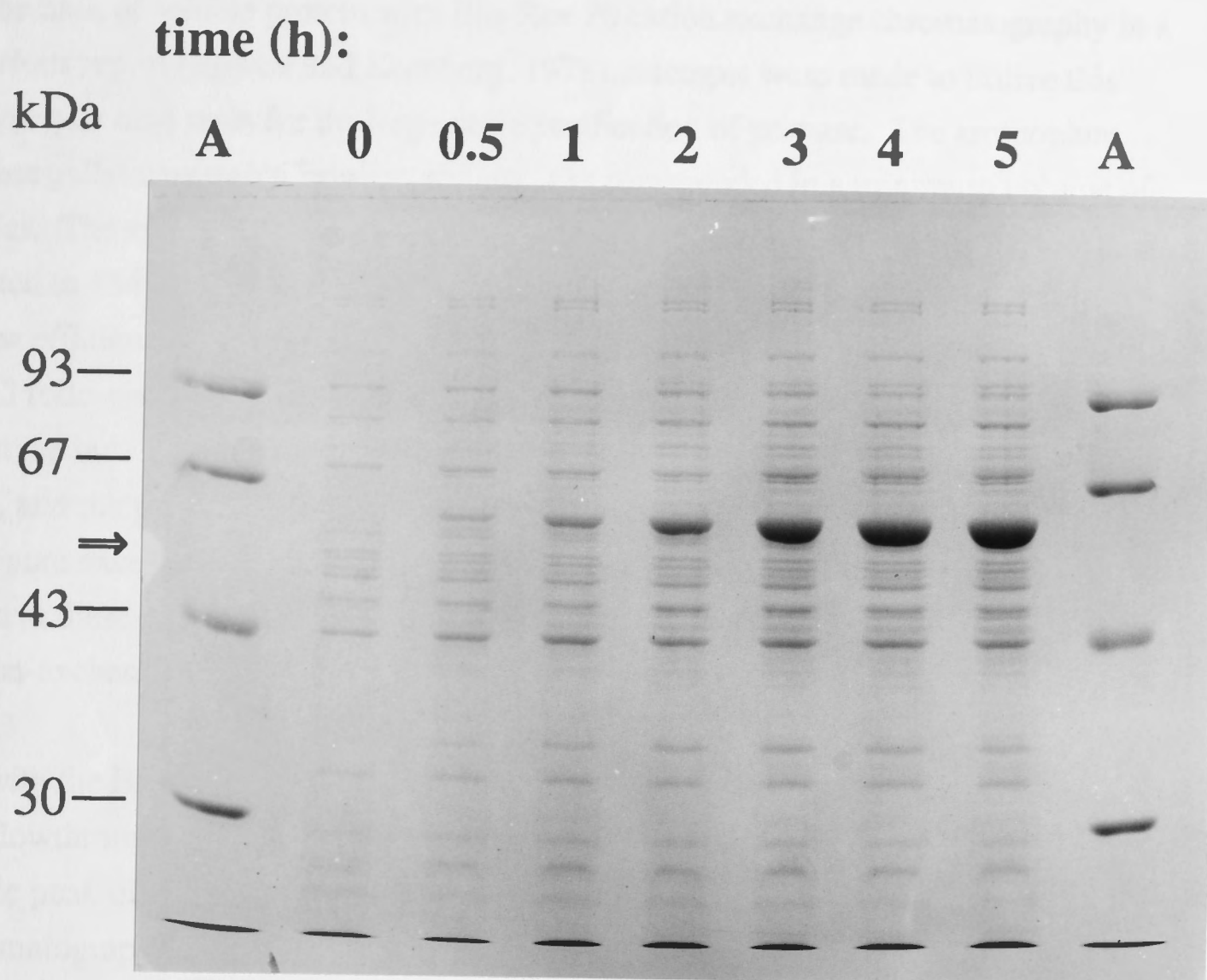


Figure 3.7

Overproduction of primase in strain RSC310. A culture (30 ml) containing RSC310 was grown at 30 °C to $A_{595} \sim 0.5$. Cells were harvested from a portion (1 ml), and the remainder of the culture was shaken at 42 °C. Cells from the 1-ml sample taken prior to commencement of treatment at 42 °C and cells from 1-ml samples taken at the indicated times thereafter were resuspended directly to $A_{595} = 10$ in SDS-gel loading buffer and treated for 2 min at 95 °C. Proteins in 10-ml were separated by 9-% polyacrylamide gel electrophoresis in the presence of SDS. Protein size markers (lane A) were as in Section 2.4.3. Following commencement of treatment at 42 °C protein accumulated with time for ~4 h by which time cell growth had ceased.

Following the successful resolution of primase from other replication proteins as well as the bulk of soluble protein with Bio-Rex 70 cation exchange chromatography in a previous report (Rowen and Kornberg, 1978), attempts were made to utilize this polyacrylic acid resin for the large-scale purification of primase. The ammonium sulfate pellet containing primase activity was resuspended in a minimum volume of buffer. The residual ammonium sulfate was removed by dialysis and the solution was diluted to 154 ml (Fraction II), effectively reducing the conductivity of the sample to allow efficient binding to the resin. Primase activity eluted between 0.35 and 0.7 M NaCl following the application of the linear NaCl gradient. This peak was separate from the bulk of the remaining soluble protein which eluted in the flowthrough (Figure 3.8), and the primase pool following this step (Fraction No's 50 to 58, Fraction III) was pure except for traces of impurities detectable by SDS-PAGE (Figure 3.10). To remove these minor contaminants, Fraction III was further fractionated by a second cation-exchange chromatography step using DEAE-Sephacel resin.

As with the Bio-Rex 70 chromatography there was no detectable primase activity in the flowthrough and all primase activity eluted in a single peak coincident with the single peak of protein between 0.14 and 0.2 M NaCl (Figure 3.9). DEAE-Sephacel chromatography improved the purity of primase only 1.01-fold, suggesting that primase was near to homogenous (Table 3.1). Primase-active fractions pooled from this step (Fraction No's 60 to 65, Fraction IV) were frozen at -70 °C and still retained full activity after a period of five months.

Properties of dnaG primase

Systems designed to express high levels of proteins from cloned genes in *E. coli* often produce protein which forms proteinaceous aggregates (Kane and Hartley, 1988). To establish that primase, when purified from a strain that overexpresses *dnaG* to such high levels, was not isolated as an aggregated species, the apparent molecular weight of the protein was examined by gel filtration chromatography. Primase eluted from a gel filtration column in the position expected for a monomeric 65-kDa protein (Figure 3.11), as observed for primase isolated from less enriched sources (Rowen and Kornberg, 1978). The suggestion has been made, based on the stoichiometry of association of primase with the G4 complementary strand origin (Stayton and Kornberg, 1983), that it may function as a dimeric species. At the concentrations used here, considerably higher than used in previous studies, there was no evidence of the existence of a dimer in the absence of single-stranded DNA.

Figure 3.8

Chromatography of primase (Fraction II) on Bio-Rex 70. Primase (Fraction II) was diluted to 154 ml with Buffer A, and applied to a column (2.5 x 12 cm) of Bio-Rex 70 cation-exchange resin pre-equilibrated in Buffer A (120 ml). Bound protein was eluted with a linear gradient (500 ml) of 0 to 3 M NaCl in Buffer A at a flow rate of 45 ml/h. Fractions (7 ml each) were collected and assayed for primase activity and for protein. Primase remained completely bound to the resin until the application of the NaCl gradient. Primase activity eluted in a single peak between 0.35 and 0.7 M NaCl.

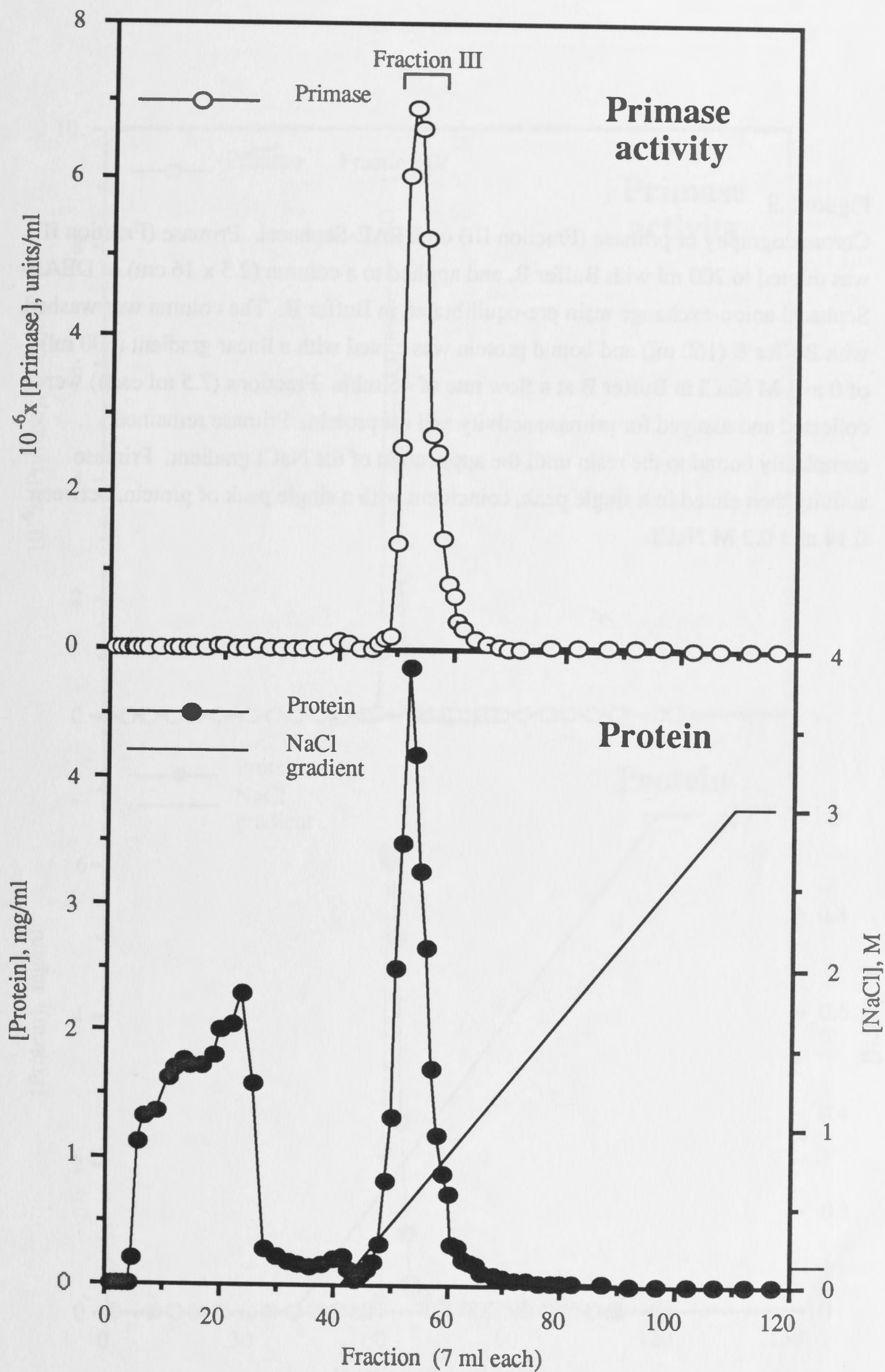
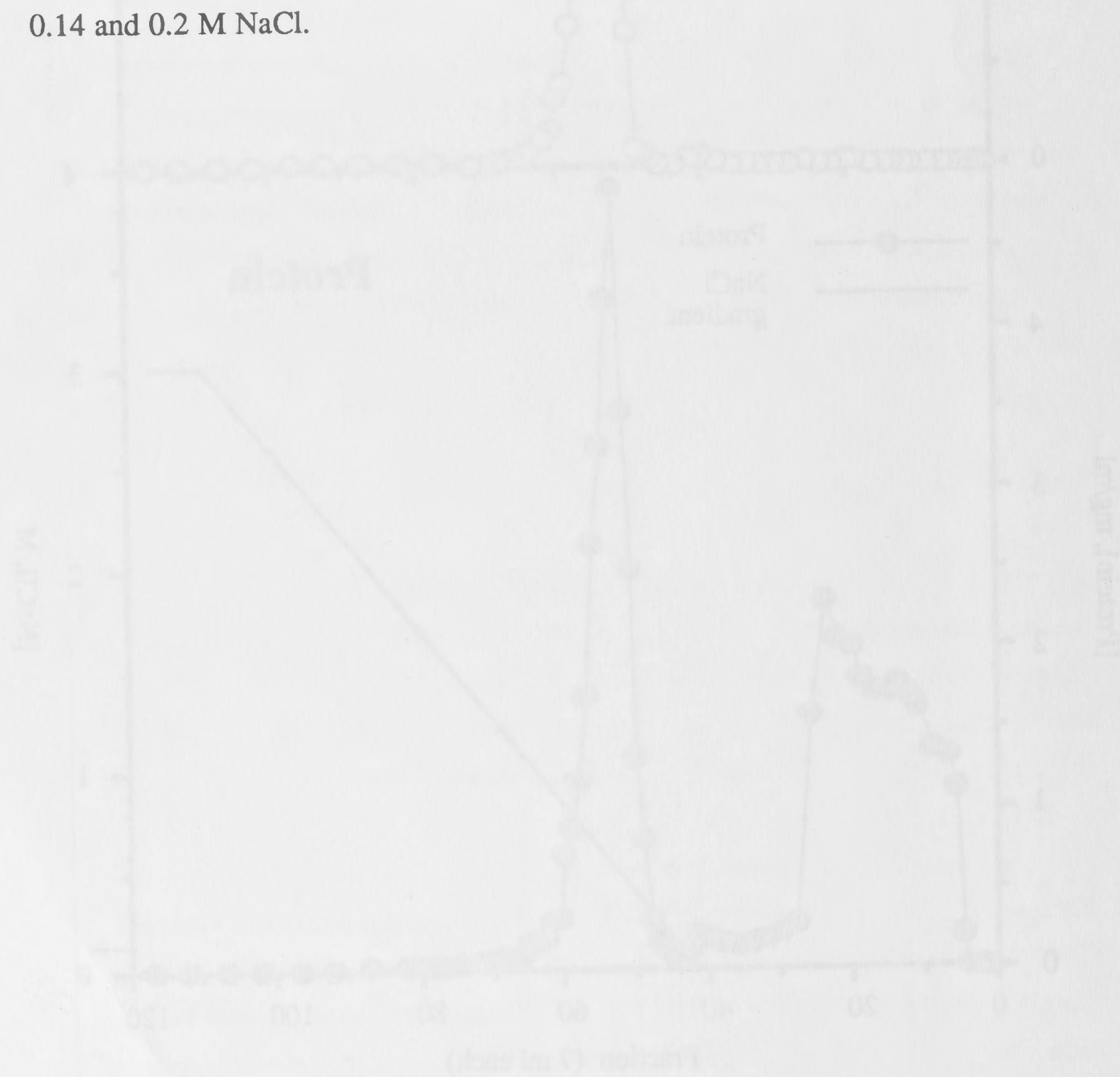


Figure 3.9

Chromatography of primase (Fraction III) on DEAE-Sephacel. Primase (Fraction III) was diluted to 200 ml with Buffer B, and applied to a column (2.5 x 16 cm) of DEAE-Sephacel anion-exchange resin pre-equilibrated in Buffer B. The column was washed with Buffer B (160 ml) and bound protein was eluted with a linear gradient (800 ml) of 0 to 1 M NaCl in Buffer B at a flow rate of 45 ml/h. Fractions (7.5 ml each) were collected and assayed for primase activity and for protein. Primase remained completely bound to the resin until the application of the NaCl gradient. Primase activity then eluted in a single peak, coincident with a single peak of protein, between 0.14 and 0.2 M NaCl.



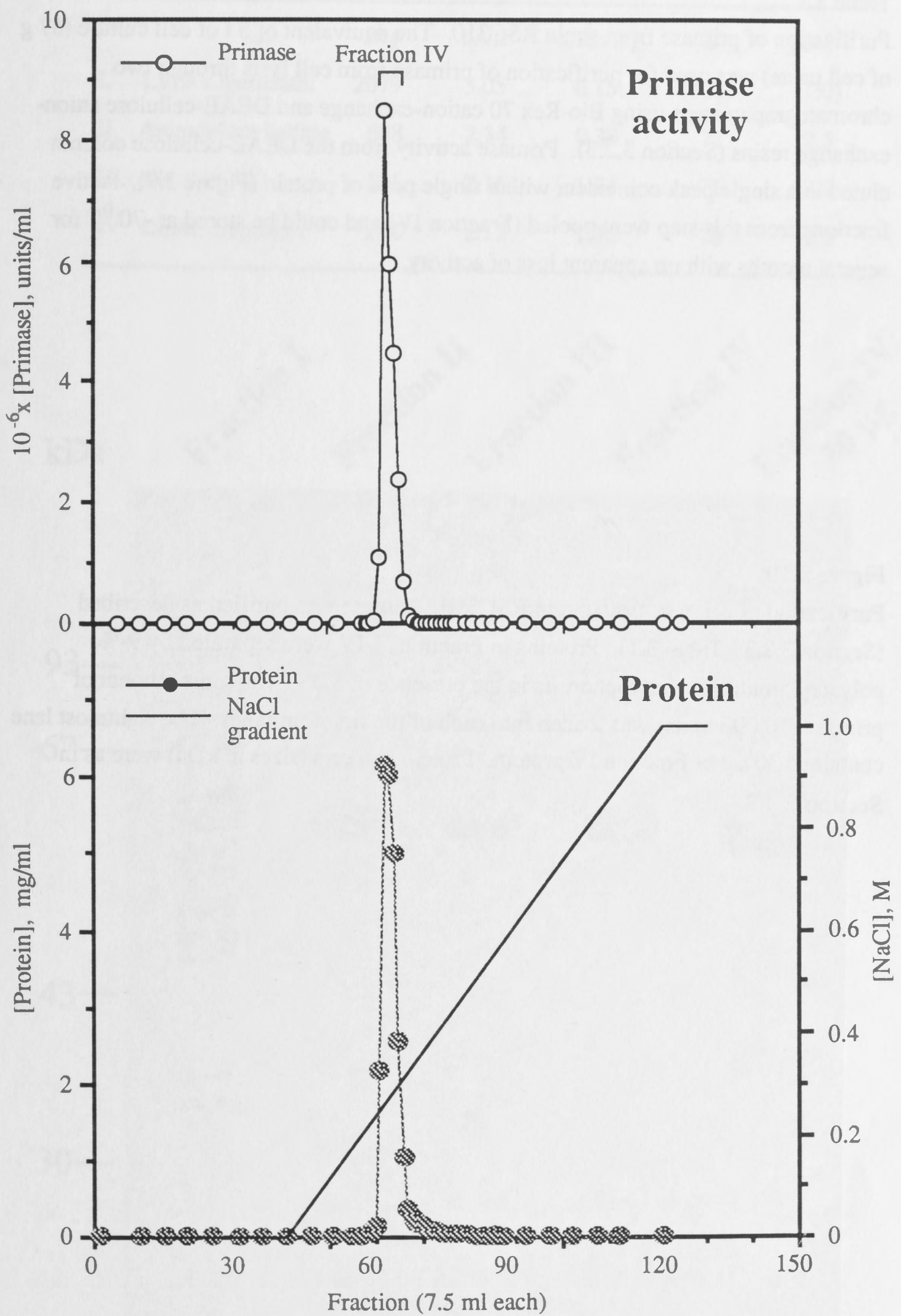


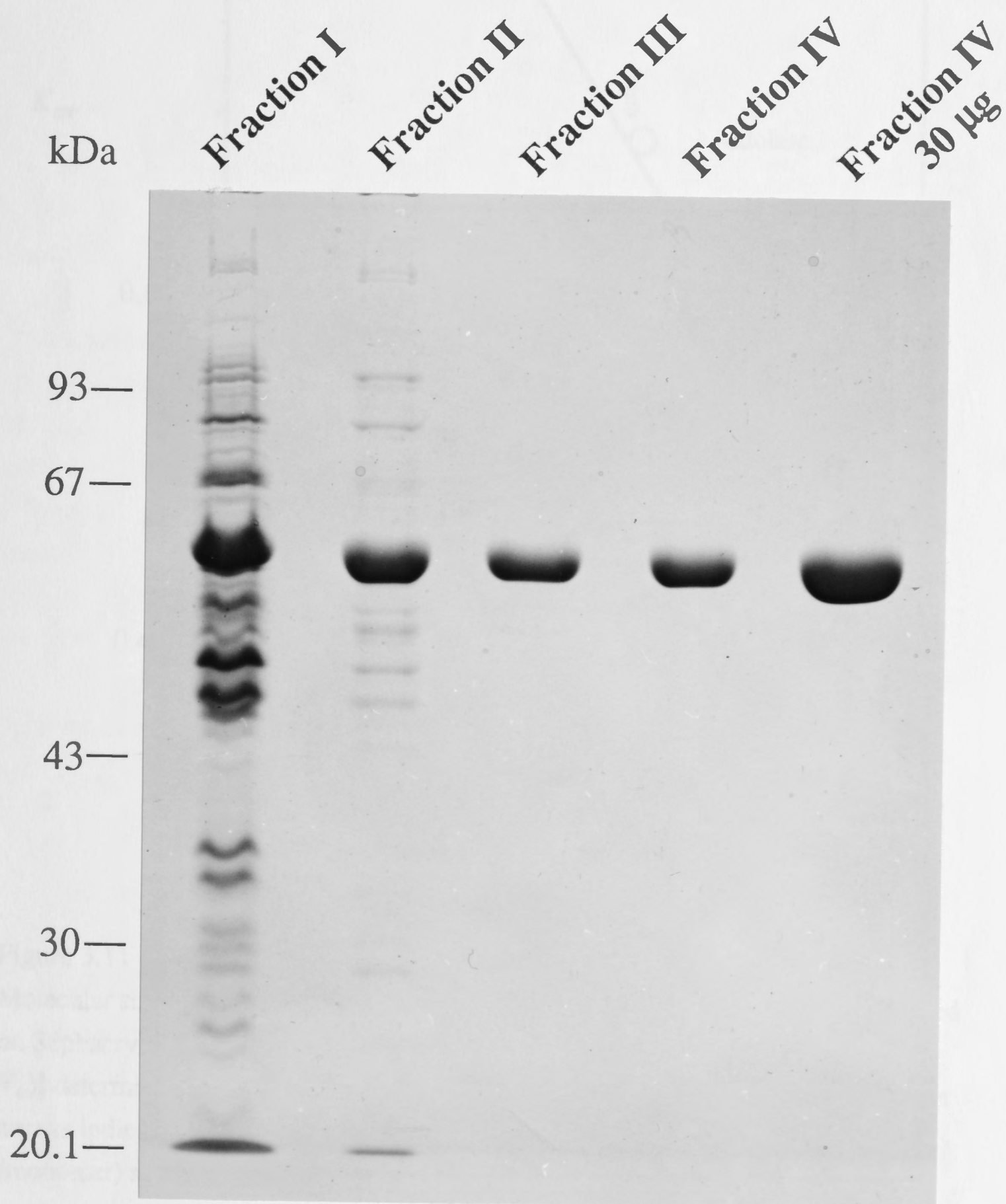
Table 3.1

Purification of primase from strain RSC310. The equivalent of 5 l of cell culture (57 g of cell paste) was used for purification of primase from cell lysis through two chromatography steps using Bio-Rex 70 cation-exchange and DEAE-cellulose anion-exchange resins (Section 3.2.3). Primase activity from the DEAE-cellulose column eluted in a single peak coincident with a single peak of protein (Figure 3.9). Active fractions from this step were pooled (Fraction IV) and could be stored at -70°C for several months with no apparent loss of activity.

Figure 3.10

Purification of primase from strain RSC310. Primase was purified as described (Section 3.2.3; Table 3.1). Proteins in Fractions I-IV were separated by 9-% polyacrylamide gel electrophoresis in the presence of SDS. The same amount of primase (10,000 units) was loaded into each of the first four lanes. The rightmost lane contained 30 μg of Fraction IV protein. Protein markers (sizes in kDa) were as in Section 2.4.3.

Fraction	Protein	$10^{-8} \times$ Activity	$10^{-6} \times$ Specific Activity	Yield	Purification
	<i>mg</i>	<i>units</i>	<i>units/mg</i>	<i>%</i>	<i>-fold</i>
I. Lysis supernatant	2079	3.05	0.15	(100)	(1.0)
II. Ammonium sulfate	608	2.34	0.38	77	2.5
III. Bio-Rex 70	161	2.16	1.34	71	8.9
IV. DEAE-Sephacel	156	2.12	1.36	70	9.1



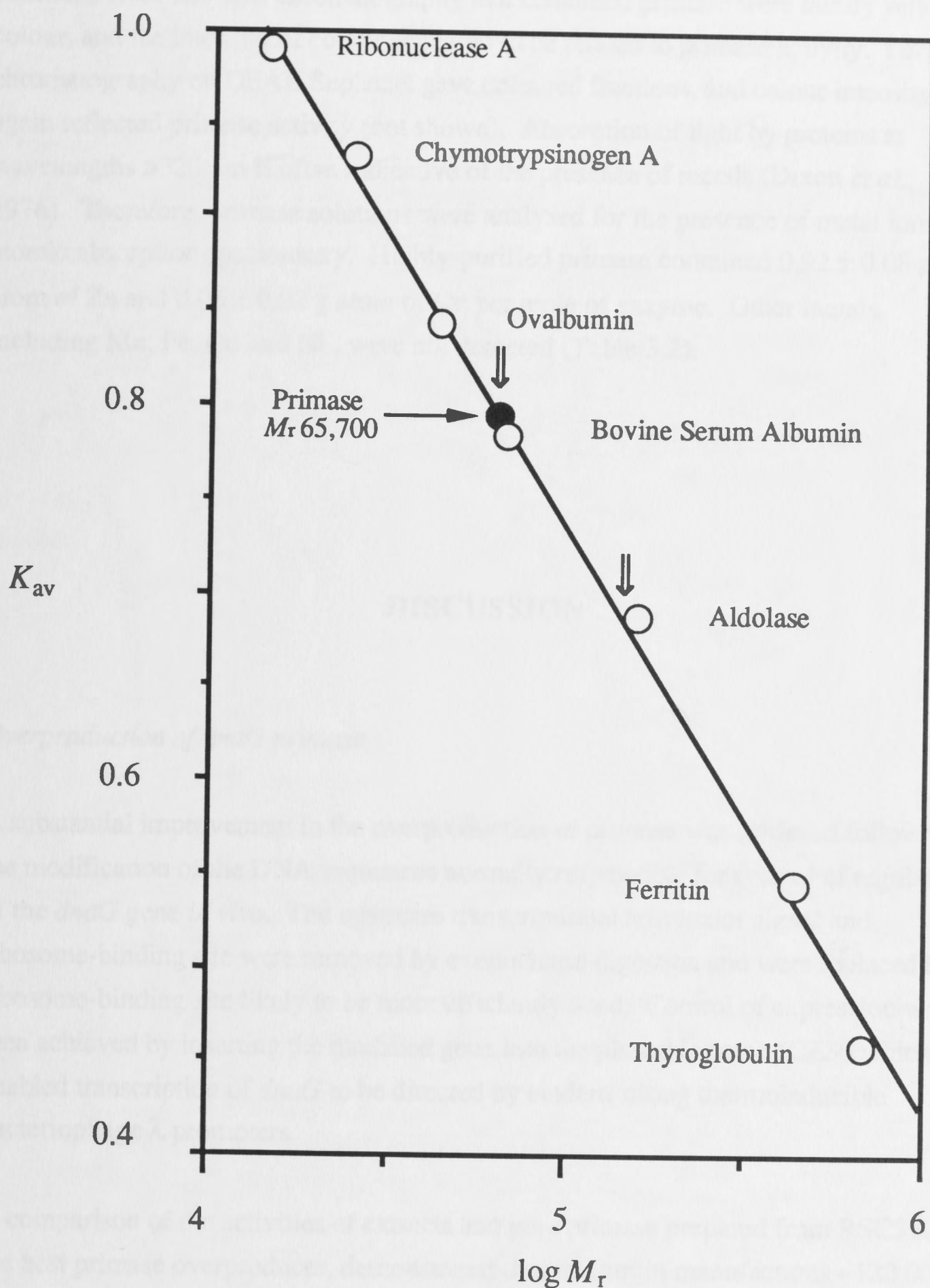


Figure 3.11

Molecular size of primase estimated by gel filtration. Proteins were chromatographed on Sephacryl S-400 as described in Section 2.4.5. The K_{av} values $[=(V_e - V_o)/(V_t - V_o)]$ determined for primase and the standards (Section 2.4.5) are plotted. The open arrows indicate the predicted K_{av} values for species of molecular mass 65.5 kDa (monomer) and 131 kDa (dimer).

Fractions from Bio-Rex chromatography that contained primase were faintly yellow in colour, and the intensity of colour appeared to be related to primase activity. Further chromatography on DEAE-Sephacel gave coloured fractions, and colour intensity again reflected primase activity (not shown). Absorption of light by proteins at wavelengths >320 nm is often indicative of the presence of metals (Dixon *et al.*, 1976). Therefore, primase solutions were analysed for the presence of metal ions by atomic absorption spectrometry. Highly-purified primase contained 0.92 ± 0.08 g atom of Zn and 0.06 ± 0.02 g atom of Cu per mole of enzyme. Other metals, including Mn, Fe, Co and Ni, were not detected (Table 3.2).

DISCUSSION

Overproduction of dnaG primase

A substantial improvement in the overproduction of primase was achieved following the modification of the DNA sequences normally responsible for control of regulation of the *dnaG* gene *in vivo*. The upstream transcriptional terminator signal and ribosome-binding site were removed by exonuclease digestion and were replaced by a ribosome-binding site likely to be more efficiently used. Control of expression was then achieved by inserting the modified gene into the plasmid vector pCE30 which enabled transcription of *dnaG* to be directed by tandem strong thermoinducible bacteriophage λ promoters.

A comparison of the activities of extracts and pure primase prepared from RSC310, the best primase overproducer, demonstrated that this strain manufactures $\sim 120,000$ molecules of primase per cell⁵, *i.e.* equivalent to levels 2,400-fold higher than those of wild-type *E. coli*. Overproduction of primase to levels that may approach these results has been recently reported (Godson, 1991), and similarly required the removal of the poor natural ribosome-binding site of the *dnaG* gene (Figure 3.3) and the upstream

5. The number of enzyme molecules per cell is an estimate based on the formula: $10^9 \times (X)/(Y)(Z)$, where X is the number of enzyme units per g of cell paste, Y the number of units per mg of pure enzyme, Z the molecular weight of the enzyme. It is estimated that 1 g of *E. coli* cell paste ($A_{595} = 400$) contains 6×10^{11} cells (Kornberg, 1980)

Table 3.2

The metal ion content of highly-purified primase samples (10 ml) from the preparation described in Section 3.2.3 (Fraction IV). Three separate samples were dialysed (Section 2.4.6), and primase activities were determined as described (Section 3.2.2). Protein concentrations were determined spectrophotometrically, using a molecular weight of 65,000 and $\epsilon_{280} = 43,240 \text{ M}^{-1}\text{cm}^{-1}$ (estimated from the amino acid composition; Gill and von Hippel, 1989) for pure primase. Values were ~3 % higher than those using the assay of Bradford (1976). Metal ions were determined by atomic absorption spectrometry as described (Section 2.4.6) using flame atomization. Analysis of Sample I for other metal ions (Mn, Fe, Co and Ni) gave levels below the limits of detection (<0.02 g atom/mol of primase).

Sample	$10^{-6} \times$ Specific Activity	[Primase]	Metal Ion	[Metal Ion]	$\frac{[\text{Metal Ion}]}{[\text{Primase}]}$
	<i>units/mg</i>	μM		μM	<i>g atom/mol</i>
I	1.32	32.4	Zn	28.5	0.88
			Cu	2.5	0.08
II	1.29	31.6	Zn	27.8	0.88
			Cu	1.4	0.04
III	1.32	50.9	Zn	51.2	1.01
			Cu	2.5	0.05

transcriptional terminator. This was achieved using the polymerase chain reaction (PCR) to reconstruct the *dnaG* coding region prior to inserting the fragment into a T7 promoter-expression vector (Studier *et al.*, 1990). In that work, primase was not purified to homogeneity nor was it analyzed for metal ions (Godson, 1991).

Apart from control of expression of the *dnaG* gene *in vivo* at the level of translation by the natural ribosome-binding site, which is a poor match to the complement of the sequence at the 3'-OH end of 16-S rRNA, some have proposed regulation of translation by the occurrence of rare codons within the coding sequence (Konigsberg and Godson, 1983). Rare codons are avoided in highly expressed genes such as those for the protein synthesis factors and ribosomal proteins. On the other hand, repressor and DNA replication genes do not discriminate against rare codons (Grosjean and Fiers, 1982). The *dnaG* gene contains an unusually high number of codons that are not frequently used in most *E. coli* genes. Translational modulation, using isoaccepting tRNA availability, may therefore form part of the mechanism regulating *dnaG* gene expression. There are now at least two clear examples of the effects of insufficiency of tRNA on expression of genes (Misra and Reeves, 1985; Brinkmann *et al.*, 1989). In each case, expression was affected by overproduction of a tRNA^{Arg}_{AGA}. The *E. coli dnaY* gene (*argU*) encodes the arginine tRNA that recognizes AGA, and possibly also AGG, codons that occur infrequently in *E. coli* genes (Garcia *et al.*, 1986). The DNA replication temperature-sensitive phenotype of *dnaY*^{ts} mutants (Henson *et al.*, 1979) suggests that expression of one or more replication genes may be sensitive to lack of the tRNA^{Arg}_{AGA}. Since the *dnaG* gene contains an AGA and AGG codon (Konigsberg and Godson, 1983) it was not possible to ignore the effect of *dnaY* gene product deficiency on overproduction of active primase. The occurrence of these rare codons within the coding sequence, however, does not appear to compromise high-level overexpression of *dnaG* (Figure 3.5). Nevertheless, the levels of expression reported here are less than those achieved by similar manipulations of other *E. coli* replication genes (Chapters 4 and 5; N.E. Dixon, personal communication), and this is perhaps due to the codon bias in *dnaG* (*cf.* Andersson and Kurland, 1990).

Purification and properties of dnaG primase

The overproduced primase from strain RSC310 (pPL195) was readily extracted in soluble form with Brij-58 and was purified by a procedure similar to that used by Rowen and Kornberg (1978) in >70 % yield. After fractionation with ammonium sulfate, primase was purified in one further step by cation-exchange chromatography

on Bio-Rex 70. Traces of impurities detectable by SDS-PAGE were removed by DEAE-Sephacel anion-exchange chromatography yielding a highly-purified preparation. The purified protein behaved as a monomeric ~65-kDa protein on gel filtration and the specific activity of primase, assessed using the G4 SS→RF assay, was 1.36×10^6 units/mg. This may be compared with reported values of 1.09×10^6 (Rowen and Kornberg, 1978), 1.2×10^6 (Stayton and Kornberg, 1983) and 3×10^6 (Kaguni and Kornberg, 1984), all determined using a similar assay. In all 156 mg of highly purified primase was isolated from 5 l of cell culture (~16 mg/g cell paste) which will enable studies of the mechanism of primase action and of its structure by X-ray crystallography.

Concentrated solutions of the pure protein were faintly coloured, suggesting the presence of a metal ion or other chromophoric cofactor. Metal ion analysis confirmed that the protein contained nearly stoichiometric quantities of Zn, together with small but significant quantities of Cu. It seems unlikely that the colour of the pure enzyme is due to its coordination to Zn(II) ions, but plausible that a small fraction of the Zn sites bind Cu(II), which might give a yellow complex. This raises first the question of the adequacy of supply of zinc in culture media, and secondly, the interesting possibility that copper might readily be substituted for zinc in the enzyme. A sample of broth as nearly identical as possible to that used for growth of the cells for purification of primase contained 11 μ M Zn and 0.9 μ M Cu. It is interesting that the two metal ions are present in a similar ratio in solutions of pure primase (Table 3.2). Calculations indicate that about 6% of the Zn and 4% of the Cu available to the cells is ultimately incorporated into primase under the conditions used for its overproduction.

The role of zinc in dnaG primase

In view of the widespread occurrence of zinc-binding sites in eukaryotic transcription factors and other proteins that interact with nucleic acids (Miller *et al.*, 1985; Berg, 1986) the sequence of primase was searched for potential donor ligands. Near the amino terminus of the protein is a sequence (Figure 3.12) that resembles Zn-binding sites in the "zinc-finger" domains of other proteins (Berg, 1990a, b). It would be expected that only four of the eight potential donor residues (cysteine or histidine) coordinate Zn in a tetrahedral complex. A comparison with the very similar amino acid sequences near the NH₂-terminus of the primases of other bacteria and bacteriophages T4 and T7 suggests which residues are likely donors (Figure 3.12). Such a cluster of donor ligands is not found elsewhere in these sequences. It has been suggested that both the T4 and T7 primases are zinc metalloproteins (Bernstein and

Figure 3.12

A. Zinc-binding regions of prokaryotic and eukaryotic primases. Sources of amino acid sequences: *E. coli*, *Escherichia coli* dnaG protein (Smiley *et al.*, 1982; Burton *et al.*, 1983); *S. typhimurium*, *Salmonella typhimurium* dnaG protein (Erickson *et al.*, 1985); *B. subtilis*, *Bacillus subtilis* dnaE protein (Wang *et al.*, 1985); Phage T4 gene 61, bacteriophage T4 gene 61 protein⁶; Phage T7 gene 4, bacteriophage T7 gene 4 protein (Dunn and Studier, 1983); Mouse, mouse primase p49 subunit (Prussak *et al.*, 1989); *S. cerevisiae*, *Saccharomyces cerevisiae* primase p48 subunit (Plevani *et al.*, 1987). X_n denotes n amino acids of unspecified sequence. The boxes enclose residues whose side chains are proposed to coordinate Zn.

B. Cysteine-rich sequences found toward the amino termini of the largest subunits of various RNA polymerases. Among the largest subunits of the vaccinia virus RNA polymerase, the yeast RNA polymerases II and III, and the *Drosophila* RNA polymerase II, an homologous NH₂-terminal region includes a cluster of cysteine residues much like the equivalent region in the β' subunit of RNA polymerase from *E. coli*. This region of the β' subunit is proposed to function in the binding of the A-site Zn(II) in the prokaryotic enzyme. Such regions might also function in zinc binding in the multisubunit eukaryotic polymerases via the two conserved Cys-X-X-Cys sequences. X_n denotes n amino acids of unspecified sequence.

Figure 3.12B taken from Coleman and Giedroc (1989).

6. M. Nakanishi and B. Alberts, unpublished (Genbank Accession No. K03113).

A.

<i>E. coli</i>	X36	-	H	A	C	C	P	F	H	-	X14	-	F	Y	H	C	F	G	C	G	A	H	-	X514
<i>S. typhimurium</i>	X36	-	H	A	C	C	P	F	H	-	X14	-	F	Y	H	C	F	G	C	G	A	H	-	X514
<i>B. subtilis</i>	X36	-	F	G	L	C	P	F	H	-	X14	-	I	F	H	C	F	G	C	G	A	G	-	X539
Phage T4 gene 61	X33	-	R	F	R	C	P	V	C	-	X21	-	N	I	H	C	Y	N	C	N	Y	H	-	X381
Phage T7 gene 4	X13	-	H	I	P	C	D	N	C	-	X12	-	H	T	F	C	Y	V	C	E	K	W	-	X527
Mouse	X127	-	A	D	I	C	S	K	C	-	X34	-	G	V	H	C	W	V	C	D	E	S	-	X347
<i>S. cerevisiae</i>	X130	-	A	Q	V	C	S	K	C	-	X34	-	G	A	H	C	W	V	S	D	K	R	-	X236

B.

<i>E. coli</i> (β')	X58	-	C	-	X11	-	C	-	X	-	C	-	X12	-	C	-	X2	-	C	-	X88
Vaccinia	X48	-	C	-	X2	-	C	-	X6	-	C	-	X30	-	C	-	X2	-	C	-	X92
Yeast pol (II)	X67	-	C	-	X2	-	C	-	X6	-	C	-	X29	-	C	-	X2	-	C	-	X110
Yeast pol (III)	X67	-	C	-	X2	-	C	-	X6	-	C	-	X29	-	C	-	X2	-	C	-	X110
<i>Drosophila</i> pol (II)	X66	-	C	-	X2	-	C	-	X6	-	C	-	X29	-	C	-	X2	-	C	-	X109

Richardson, 1988). Although this proposal has yet to be tested experimentally, the present analysis of the *E. coli* protein strengthens the case for the presence of Zn in all of the known primases, extending even to the eukaryotic enzymes (Figure 3.12) that also contain similar sequences near their amino termini (Prussak *et al.*, 1989).

Fortuitously, a strain that overproduced an NH₂-terminal deletion mutant of primase (primase- Δ 27) was isolated following digestion of *dnaG* by nuclease Bal31 during the original cloning procedures (Figure 3.3). Although primase overproduced from this source was at least partly soluble and presumably therefore, folded into a stable conformation, it was completely inactive in priming G4 phage DNA replication *in vitro*. The deletion (27 residues) was not sufficient to remove any of the proposed zinc ligands. It seems plausible, however, that the Zn-binding domain may have been disrupted. By analogy with the transcription factors, Bernstein and Richardson (1988) have suggested that the region in T7 primase proposed to bind Zn is also involved in binding to DNA. The NH₂-terminal region of the *E. coli* primase (residues 1 to 36) that immediately precedes the putative Zn-binding site contains four arginine and four lysine residues, and could therefore participate in DNA binding. The first four of these basic residues are absent from primase- Δ 27. While a role for the NH₂-terminal region in binding DNA is an attractive proposition, there are reasons for exercising caution. As examples, both the T4 gene 32 single-stranded DNA-binding protein and the SV40 large T antigen contain Zn coordinated in (putative) zinc finger structures, but the regions of both proteins that contain the metal ions are more involved in protein-protein interactions than in DNA binding (Berg, 1990b; Giedroc *et al.*, 1987; Loeber *et al.*, 1991). It also seems plausible that this motif will recur in structures of many metalloproteins, regardless of function, simply because it is a convenient way to satisfy the preference of Zn for thiolate (or histidine) ligands in a tetrahedral coordination geometry.

The primases join a growing family of Zn-containing proteins that bind without strict sequence specificity to DNA. *E. coli* proteins in this class include the *uvrA* protein (Navaratnam *et al.*, 1989), DNA topoisomerase I (Tse-Dinh and Beran-Steed, 1988), and most likely also the *priA* protein (Nurse *et al.*, 1990; Lee *et al.*, 1990). Although recognition of DNA by primase, *uvrA* and *priA* is not sequence specific, each is thought to bind particular structural features in DNA. The *uvrA* protein recognizes distortions in double-stranded DNA at lesions produced by UV damage and chemical reagents (Thomas *et al.*, 1986; van Houten, 1990). The *priA* protein recognizes secondary structural elements in single-stranded DNA at the complementary strand origin of phage ϕ X174 DNA and at other primosome-assembly sites (Masai *et al.*,

1990a), while primase recognizes directly the hairpin structures at the phage G4 complementary strand origin (Stayton and Kornberg, 1983).

Insight into structural or catalytic roles for Zn(II) in primase function may best be gained by reviewing the role of zinc in the RNA and DNA polymerases. Initially it appeared that DNA and RNA polymerases were universally zinc metalloenzymes (Wu and Wu, 1983; Coleman, 1983), although subsequently it has been shown that DNA polymerase I from *E. coli* (Ferrin *et al.*, 1983) and the phage T7 DNA polymerase (Slaby *et al.*, 1984) do not contain zinc. Clearly then, the basic enzymatic polymerization of ribo- and deoxyribonucleotides does not necessarily require zinc as a catalytic cofactor.

The RNA polymerase from *E. coli*, a holoenzyme consisting of five subunits $\alpha_2\beta\beta'\sigma$, contains two tightly incorporated Zn(II) ions per molecule (Giedroc and Coleman, 1986). One is tightly bound in the β' subunit (the A-site), while the other is less tightly bound and may be present at or near the interface between the β and β' subunits (the B-site). The precise function (either structural or catalytic) of the two intrinsic Zn(II) ions in the transcription process by RNA polymerase are not well defined although the B-site Zn(II) does not function directly in the catalytic reaction or binding of the template. The A-site zinc is bound in a tetrahedral-cysteine thiolate complex located at the amino terminus of the protein (Coleman and Giedroc, 1986; Solaiman and Wu, 1985). Similar sequences near the amino terminus of other Zn-containing RNA polymerases have also been identified (Figure 3.12). These eukaryotic RNA polymerases, like the bacterial enzyme, have a common molecular architecture consisting of two large subunits conserved during evolution (Bartholomew *et al.*, 1986). Zinc might therefore be envisaged to play at least some part in the chemistry performed by these enzymes.

Nucleic acid chemistry is largely phosphate ester chemistry, and there is a wealth of evidence from chemistry and enzymology for metal ion promotion of reactions of phosphate esters, where the role of the metal ion is as a Lewis acid active at neutral pH (Dixon and Sargeson, 1983; Hendry and Sargeson, 1990). It would not be too surprising therefore if it were found that the zinc ion in primase and these polymerases is involved in polynucleotide synthesis at the active site of the enzyme rather than solely in its binding to DNA. It has been suggested that Zn(II) participates in catalysis by coordinating the intrinsic metal with the 3'-OH of the ribonucleotide, thereby facilitating its deprotonation (Mildvan, 1977), a necessary step for nucleophilic attack of the α -phosphate of the incoming nucleotide. An alternative hypothesis involves coordination of the Zn(II) with the 2'-OH (Wu and Wu, 1983) of the initiating

nucleotide and this may serve as a basis for discrimination between dNTPs and rNTP substrates by RNA polymerase. In this context, it is interesting that primase can substitute dNTPs for rNTP in all but the first, and possibly the second, position in a primer to produce hybrid primers interspersed with ribo- and deoxyribonucleotides (Rowen and Kornberg, 1978). The specific requirement of Mg^{2+} in the primase-catalysed reaction (Stayton and Kornberg, 1983) may indicate the involvement of Mg^{2+} in coordinating the incoming nucleotide prior to catalysis.

Other than binding the substrate by direct coordination with the initiating nucleotide other roles proposed for Zn(II) in *E. coli* RNA polymerase include interactions with the DNA template and a structural role in the maintenance of the active conformation of the enzyme (Wu, 1986). However, the single subunit RNA polymerase encoded by bacteriophage T7 contains no intrinsic zinc (King *et al.*, 1986). If the phage and bacterial RNA polymerases can be considered to share a common catalytic mechanism, then the T7 enzyme indicates that zinc need not perform a catalytic role. The role of Zn in the RNA polymerase of *E. coli* therefore requires further definition. Although electron diffraction has been used to obtain structural information at lower resolution (Darst *et al.*, 1989), it is unfortunate that high-resolution structural information of the large multi-subunit RNA polymerases like *E. coli* RNA polymerase has not yet been achieved. The isolation of large quantities of the single polypeptide of primase from *E. coli* may therefore present a unique opportunity to investigate the role of zinc in both this enzyme and in a broader spectrum of related proteins.

The bulk of evidence supporting the proposals for the role of zinc in RNA polymerase has come from analysis of the enzyme following substitution of the intrinsic metal ion (reviewed in Coleman and Giedroc, 1989; Wu, 1986). Direct proof of the functional role of metal ions in a metalloenzyme is the removal and readdition of the intrinsic metal with the concurrent loss and restoration of enzymatic activity. Such an approach has not been readily possible with RNA polymerase, although *in vivo* substitution of zinc with cobalt(II) by growing *E. coli* cells in Zn-depleted and Co-enriched media (Speckhard *et al.*, 1977) has made such investigations feasible. These experiments may also be possible with the *E. coli* primase if incorporation of Cu(II) or other metals can similarly be effected by *in vivo* substitution or metal ion depletion and reconstitution *in vitro*. Investigations of this nature and site-directed mutagenesis aimed at identifying the donor ligands involved in coordinating the intrinsic metal ion are currently being carried out in conjunction with the crystallization and structural determination of the enzyme.

CHAPTER 4

Improved Overproduction of the dnaB, dnaC and dnaT Proteins

Chain elongation by DNA polymerase III holoenzyme requires RNA priming, a reaction strictly dependent on the action of primase to synthesize the primer fragments. Although primase is able to prime DNA replication independently on SSB-coated G4 single-stranded DNA, primer formation on other DNA templates and DNA templates not coated with SSB requires at least *dnaB*, and often other primosomal proteins. With the focus of this project on the study of protein-protein and protein-DNA interactions in the primosome, isolation of the other primosomal proteins in quantities sufficient for structural and chemical analysis was a priority. As for primase, expression of these proteins are all tightly regulated *in vivo* and current vectors designed for their overproduction only produce quantities sufficient for enzymological studies. Following success in overproducing the *dnaG* gene product, additional experiments were undertaken to overproduce the *dnaT*, *dnaC* and *dnaB* primosomal proteins whose genes were currently available.

The *dnaT* and *dnaC* genes are transcriptionally coupled and form an operon together with two other uncharacterized genes (Masai and Arai, 1988b; Figure 4.1). The regulation of these genes is thought to be mediated by translational control and mRNA processing. The operon, located at 99 minutes on the *E. coli* chromosomal map, has two major transcriptional termination sites downstream of the *dnaC* gene and both *dnaT* and *dnaC* are transcribed from a promoter located 104 bp upstream of *dnaT*.

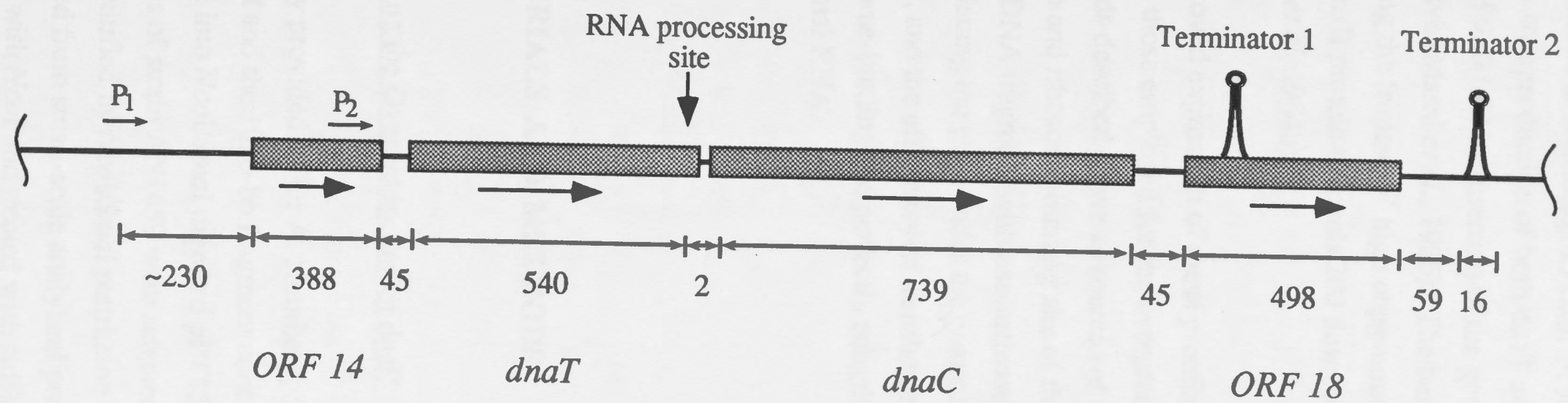
Steady state levels of these proteins in wild-type *E. coli* have been determined at 50 molecules of *dnaT* protein (Arai *et al.*, 1981g) and 100 molecules of *dnaC* protein (Kornberg, 1982) per cell. This two fold excess of *dnaC* over *dnaT* protein is proposed to result from a combination of a higher percentage of rare codons in the *dnaT* gene, more efficient translation initiation on the *dnaC* mRNA due to a superior ribosome-binding site, and post-transcriptional modification of *dnaT* mRNA (Masai and Arai, 1988b). Evidence suggests that the *dnaT-dnaC* transcript is cleaved immediately upstream of the *dnaC* gene and the processed mRNA containing the *dnaT* gene appears to be degraded rapidly from the 3' end. This bears some analogy to control of expression of *dnaG* (Section 3.1).

That expression of the *dnaB* gene is highly regulated *in vivo* is suggested by its low cellular abundance. There are estimated to be about 20 molecules of *dnaB* protein per cell in wild-type *E. coli* (Ueda *et al.*, 1978). The *dnaB* gene has been isolated and maps at 92 minutes on the *E. coli* chromosomal map. The coding region of *dnaB* (Nakayama *et al.*, 1984b) lacks a typical ribosome-binding site, and this may account for its low level of expression. No other regulatory features have been identified nor has the immediate region surrounding the *dnaB* gene been sequenced.

Figure 4.1

Arrangement of the regulatory units of the *E. coli dnaT-dnaC* operon. Arrows indicate the direction of transcription and the numbers below the line represent in base pairs the sizes of the genes and regions between the genes and regulatory units. The *dnaT* and *dnaC* genes are cotranscribed from a promoter (P2) located upstream of the *dnaT*. There is evidence that transcription is also initiated from a second promoter further upstream, although the contribution to overall levels of *dnaT* and *dnaC* proteins in the cell from these transcripts is negligible. Some of the transcripts extending through the *dnaC* gene are terminated at a terminator immediately downstream of the *dnaC* gene. The majority, however, pass through this terminator and are terminated downstream of *ORF-18*. Further regulation of the *dnaT* gene is achieved by the processing of a portion of the transcripts at two sites proximal to *dnaC* and located within the *dnaT* coding region.

Figure adapted from Masai and Arai (1988b).



All three genes have been cloned and overexpressed in *E. coli* with varying degrees of success. Overproduction of the *dnaC* protein was achieved by increasing gene dosage in the cell. The *dnaC* gene was inserted into the temperature-inducible high copy-number plasmid pMOB45 (Kobori and Kornberg, 1982a). A similar method of gene amplification was also used for overproduction of both *dnaT* and *dnaB* proteins. The *dnaT* protein was overproduced >20 fold by insertion of the gene into the polylinker region of a pUC vector derivative (Masai *et al.*, 1986). The best overproduction of *dnaB* protein was achieved using the 'runaway' high copy-number plasmid pBEU17, resulting in expression of the *dnaB* protein to levels 200 times that of the host strain (Arai *et al.*, 1981b; Nakayama *et al.*, 1984b).

In the work reported here, improved expression of these proteins was achieved by following procedures similar to those employed for the overproduction of primase (Chapter 3) utilizing the plasmids described above as sources of the genes. In each case the natural promoter region and ribosome-binding site of the gene were removed by treatment of the appropriate DNA fragment with exonuclease Bal31. Improved transcription was achieved by placing the gene under the control of tandem λ promoters in pCE30 derivatives, and the efficiency of translation was increased by providing the gene with a ribosome-binding site perfectly complementary to the 3' terminus of *E. coli* 16-S ribosomal RNA.

MATERIALS AND METHODS

4.2.1 A Plasmid Directing the Overproduction of *dnaC* Protein

The plasmid pJK129 (generously provided by Dr A. Kornberg, Stanford University) was digested with *Nco*I and *Hpa*I and the 1920-bp fragment containing the intact *dnaC* gene was isolated and subcloned into *Nco*I/*Sma*I digested pPT150 (Figure 4.2). Ampicillin resistant transformants of strain AN1459 were selected at 30 °C and a plasmid containing *dnaC* was identified by *Nco*I/*Sma*I restriction endonuclease digestion of plasmid DNA isolated from small-scale analytical preparations. The new plasmid, pPS231, was linearized with *Nco*I and treated with sufficient exonuclease Bal31 to remove ~288 bp of DNA upstream of the ATG translation initiation codon of the *dnaC* gene. The resultant fragment mixture was isolated, digested with *Hpa*I and

recircularized by intramolecular religation. Ampicillin-resistant transformants were selected in AN1459 and purified, then screened for temperature sensitivity at 42 °C. Selected plasmids of the anticipated size were further screened by estimation of sizes of *Bam*HI/*Mlu*I fragments. For further screening, plasmid DNA from small-scale analytical preparations was digested with *Bam*HI and *Hae*III and 3'-end labelled with Klenow polymerase and [α -³²P]dATP. Sizes of fragments were estimated by polyacrylamide gel electrophoresis and autoradiography (Section 2.3.11). Plasmids predicted to contain a short RBS-ATG spacer were then purified and sequenced using the dideoxy method. All selected strains were analysed for the expression of *dnaC* protein at 42 °C by visualization of total cell proteins on 12 % SDS-polyacrylamide gels (Figure 4.3).

4.2.2 A Plasmid Directing the Overproduction of *dnaT* Protein (Protein i)

The plasmid pHM3089 (generously provided by Dr H. Masai, DNAX Research Institute) was digested with *Cla*I and *Hind*III and the ~1150-bp fragment containing the intact *dnaT* gene was isolated. The fragment was then treated with an appropriate amount of exonuclease Bal31 to remove 100-160 bp of DNA upstream of the ATG translation initiation codon of the *dnaT* gene. The mixture of blunt-ended product fragments were then ligated into 5'-dephosphorylated *Hpa*I-linearized pND201 vector. Competent AN1459 were transformed with the resultant plasmid pool and ampicillin resistant transformants were selected at 30 °C. Plasmids of anticipated size were screened for the presence and orientation of inserts by *Bg*III/*Nco*I restriction endonuclease digest of plasmid DNA isolated from small-scale analytical preparations. The extent of exonuclease Bal31 digestion was judged by 3' end-labelling studies of plasmid DNA digested with *Bam*HI and *Hpa*II. Plasmids predicted to contain a RBS-ATG spacer of <60 bp were then purified and sequenced using the dideoxy method. From this initial cloning experiment a plasmid bearing the *dnaT* gene 50 bp downstream of the *Bam*HI site in pND201 was isolated (pPS298; Figure 4.4). This plasmid was linearized with *Bam*HI and treated with enough exonuclease Bal31 to remove the remaining DNA upstream of the ATG translation initiation codon. The resulting mixture of blunt-ended fragments was digested with *Eco*RI; the *dnaT*⁺ fragments (~1,000 bp) were isolated and ligated between the *Hpa*I and *Eco*RI sites of pPT150. The product plasmid pool was then used to transform competent AN1459 and ampicillin resistant transformants, selected at 30 °C, were screened for temperature sensitivity at 42 °C. Plasmids predicted by estimation of sizes of ³²P-end-labelled *Bam*HI/*Hpa*II fragments to contain a short RBS-ATG spacer were then purified and sequenced using the dideoxy method. All selected strains were analysed

for the expression of the *dnaT* protein at 42 °C by visualization of total cell protein on 13 % SDS-polyacrylamide gels (Figure 4.5).

4.2.3 A Plasmid Directing the Overproduction of *dnaB* Protein

The plasmid pKA1 (generously provided by Dr A. Kornberg, Stanford University) was digested with *Bam*HI and *Hind*III and an ~4.4-kbp fragment containing the intact *dnaB* and *npt* (conferring kanamycin resistance) genes was isolated and ligated between the *Bam*HI and *Hind*III sites of pUC9 (Figure 4.7). Transformants of AN1459 resistant to both ampicillin and kanamycin were selected at 30 °C and a plasmid containing *dnaB* was identified by *Bam*HI/*Hind*III restriction endonuclease digestion of plasmid DNA isolated from small-scale analytical preparations. The product, pPS307, was introduced into the *dnaB*^{ts} strain SG1692 *recA* by transformation. All transformants selected for ampicillin resistance at 30 °C were also kanamycin resistant, and grew well at 42 °C (restrictive temperature for SG1692 *recA*). Plasmid pPS307 was then digested with *Nde*I. The *dnaB*⁺ fragment (~2,800 bp) was isolated and treated with sufficient exonuclease Bal31 to remove ~44 bp of DNA upstream of the ATG translation initiation codon. The resulting blunt-ended fragments were isolated and further digested with *Eco*RI. The *dnaB*⁺ fragments (~2.1 kb) were isolated and ligated between the *Hpa*I and *Eco*RI sites of pPT150. After transforming competent AN1459 with this plasmid pool, ampicillin-resistant transformants were selected at 30 °C and screened for temperature sensitivity at 42 °C. Following this the presence and orientation of inserts and approximate end-points of Bal31 digestion were determined by *Bam*HI/*Nco*I restriction endonuclease digests of plasmid DNA isolated from small-scale analytical preparations. All selected strains were analysed for the expression of *dnaB* protein at 42 °C by visualization of protein in lysed cells using SDS-PAGE (Figure 4.8). Those plasmids that successfully directed the overproduction of *dnaB* were then purified and sequenced using the dideoxy method. The selected *dnaB*-overproducing plasmid (pPS353) was further modified by linearization of the plasmid DNA with *Sma*I and recircularization of the large fragment to give pPS359.

4.2.4 Dideoxy sequencing of pPS359

DNA downstream of the *dnaB* gene in pPS359 was sequenced directly on one strand using primer 10 (Figure 2.1). Further sequences were obtained using two other plasmid derivatives: (i) An 822-bp *Bgl*III/*Nco*I fragment from pPS359 containing the

3'-OH terminus of the *dnaB* gene was isolated and ligated between the *Bam*HI and *Nco*I sites of pMTL23P. Ampicillin resistant transformants of AN1459 were selected and a plasmid bearing the inserted fragment was identified by digestion of plasmid DNA from small-scale analytical preparations with *Eco*RV/*Nco*I. The new plasmid construct (pC-1) was purified and sequence determined using the M13 universal primer. (ii) The second sequence construct was engineered by insertion of a 378-bp *Bgl*II/*Eco*RI fragment containing the DNA sequence distal to the 3'-OH terminus of the *dnaB* gene between the *Bam*HI and *Eco*RI sites of pPS436 (Section 4.2.5.1). Ampicillin resistant transformants of AN1459 were selected and a plasmid bearing the inserted fragment was identified following digestion of plasmid DNA from small-scale analytical preparations with *Nco*I/*Eco*RI. The new plasmid construct (pC-2) was purified and sequence was determined using primer 9 (Section 2.1).

4.2.5 Concurrent Overproduction of *dnaB* and *dnaC* Proteins

4.2.5.1 Construction of the Dual Promoter Vector pPS436

The vector pCE30 was linearized with an *Rsa*I partial digest and the linear DNA was purified from other species (Figure 4.9). The linear plasmid was further digested with *Bgl*II and the 3924-bp fragment was isolated. The cloning vector pMTL20P was digested with *Hind*III and the ends were filled with Klenow fragment. The blunt-ended linear plasmid was subsequently digested with *Bam*HI and the 58-bp polylinker fragment was isolated. This fragment and the *Bgl*II/*Rsa*I digested pCE30 DNA were ligated, thus ensuring insertion of the polylinker in only one orientation. Ampicillin resistant transformants of AN1459 were selected at 30 °C. Plasmids isolated from small-scale analytical preparations were screened for the presence of newly acquired restriction sites. The orientation of the inserted region of the new plasmid vector, pPS436, was determined by DNA sequencing using as template a plasmid derivative carrying a deletion between the *Eco*RI and *Stu*I sites (using primer 10, Figure 2.1).

4.2.5.2 High-level Expression of *dnaB* and *dnaC* from pPS436 Derivatives

The *Bam*HI/*Eco*RI *dnaB*⁺ and *dnaC*⁺ cassettes from pPS359 (2118 bp; Figure 4.7) and pPS237 (1658 bp; Figure 4.2), respectively, were isolated and in separate reactions were inserted behind *P*_L in pPS436 by ligation with *Bam*HI/*Eco*RI-linearized vector (Figure 4.10). Ampicillin-resistant transformants of AN1459 were selected at 30 °C. The two new recombinant plasmids, pPS526 and pPS527, were identified by *Bam*HI/*Eco*RI restriction endonuclease digests of plasmid DNA isolated from small-

scale analytical preparations. Both successfully overproduced their respective proteins (Figure 4.13).

The *Bam*HI/*Eco*RI RBS-gene fusion cassettes previously isolated from pPS237 and pPS359 were end-filled with Klenow fragment and ligated into plasmids pPS526 and pPS527, respectively, each of which had been linearized with *Stu*I and 5'-dephosphorylated (Figure 4.10). The insertion of these fragments directed the re-formation of both *Bam*HI and *Eco*RI sites at the 3'-OH and 5'-P termini of the cassettes. Ampicillin resistant transformants of AN1459 were selected at 30 °C and plasmids harbouring both *dnaC* and *dnaB* genes were identified by *Bam*HI and *Eco*RI restriction endonuclease digests of plasmid DNA isolated from small-scale analytical preparations. These plasmids were designated (P_R -*dnaC*⁺/ P_L -*dnaB*⁺) pPS553 and (P_R -*dnaB*⁺/ P_L -*dnaC*⁺) pPS554.

4.2.5.3 Construction of a Plasmid Bearing a *dnaC*-*dnaB* Synthetic Operon

The plasmid directing high-level expression of the *dnaC* gene, pPS237 (Figure 4.2), was digested with *Acc*I and the linear DNA was end-filled with Klenow enzyme. The isolated linear DNA was then further digested with *Eco*RI. The large fragment containing all of the vector and *dnaC* gene was purified. Concurrently, the plasmid pPS359 (*dnaB*⁺) was digested with *Bam*HI and end-filled. The linearized blunt-ended plasmid was further digested with *Eco*RI and the 2120-bp fragment containing the *dnaB* gene fused to the synthetic ribosome-binding site of the parent plasmid was isolated. This fragment was then ligated with the linear vector DNA harbouring the *dnaC* gene, resulting in the fusion of the *dnaB* gene immediately downstream of the terminus of the *dnaC* gene. The product plasmid (pPS562; Figure 4.11) was identified on the basis of *Bam*HI/*Eco*RI restriction endonuclease digests of plasmid DNA isolated from small-scale analytical preparations from ampicillin resistant AN1459 transformants selected at 30 °C. Overproduction of the desired proteins at 42 °C was confirmed by SDS-PAGE.

4.2.6 Chemical Induction of High-Level Expression of the *dnaB* Protein

The 2309-bp *Eco*RI fragment from pPS553 that contains *dnaB* (Figure 4.10) was isolated and ligated into the *Eco*RI site in the vector pMTL22P (Figure 4.12). Ampicillin-resistant transformants of AR120 (*cI*⁺) were selected at 30 °C. A plasmid containing the *dnaB*⁺ fragment in the correct orientation (pPS561) was identified by digestion of plasmid DNA from small-scale analytical preparations with *Nco*I.

Chemical induction of *dnaB* overproduction was achieved by growing AR120/pPS561 at 37 °C to $A_{595} = 1.0$, followed by the addition of nalidixic acid to 40 µg/ml and incubation of the culture with aeration for a further 16 h.

RESULTS

Overproduction of dnaC protein

The overproduction of the *dnaC* protein was achieved using the plasmid vector pPT150, a derivative of pCE30 (Elvin *et al.*, 1990). This plasmid bears a ribosome-binding site perfectly complementary to the 3'-terminus of 16-S ribosomal RNA immediately upstream of a unique *HpaI* site. Unlike pCE30, the vector therefore allows direct fusion of the translation initiation codon of a gene in close proximity to the synthetic ribosome-binding site. This vector also has several advantages over the analogous plasmid, pND201. It contains a group of unique restriction endonuclease sites 350 bp downstream of the *HpaI* site and ribosome-binding site. Problems encountered in obtaining sufficient numbers of transformants in the correct orientation may therefore be eliminated by ensuring correct orientation of the gene prior to exonuclease Bal31 treatment. This feature eliminates inefficiencies incurred in ligation of blunt-ended DNA fragments and removes difficulties arising from reclosure of vector molecules that had not been properly 5'-dephosphorylated.

Insertion of the *dnaC* gene from pJK129 into pPT150 positioned the gene in the correct orientation with respect to the tandem λ promoters (Figure 4.2). Following Bal31 treatment of the product plasmid (pPS231) the synthetic ribosome-binding site of the vector was brought into proximity with the ATG translation initiation codon of *dnaC* by removal of residual DNA downstream of the ribosome-binding site with *HpaI* and recircularization of the plasmid. This strategy yielded a number of excellent overproducers of the full length *dnaC* protein and two NH₂-terminal deletion mutants (Figure 4.3).

The screening procedures employed to isolate these overproducers were both involved and time consuming. The region of DNA required to be removed to approach the ATG translation initiation codon of the *dnaC* gene was extensive (288 bp) and

Figure 4.2

Construction of a plasmid, pPS237, that directs overproduction of *dnaC* protein. Details of manipulations are described in Section 4.2.1. The *Mlu*I site used in screening of plasmids of anticipated size for the presence and orientation of inserts is centred 647 bp from the 5' end of the *dnaC* coding region. The *Hae*III sites (*) used for 3'-end labelling experiments are located 93 bp from the 5' end of the *dnaC* coding region and 1158 bp upstream of the *Bam*HI site in the vector. Although there are 12 other *Hae*III sites within pPS237, fragments from between these sites are not radioactively end-labelled and are not visible following autoradiography. The relevant sequences of plasmids selected from this experiment are given (Figure 4.3).

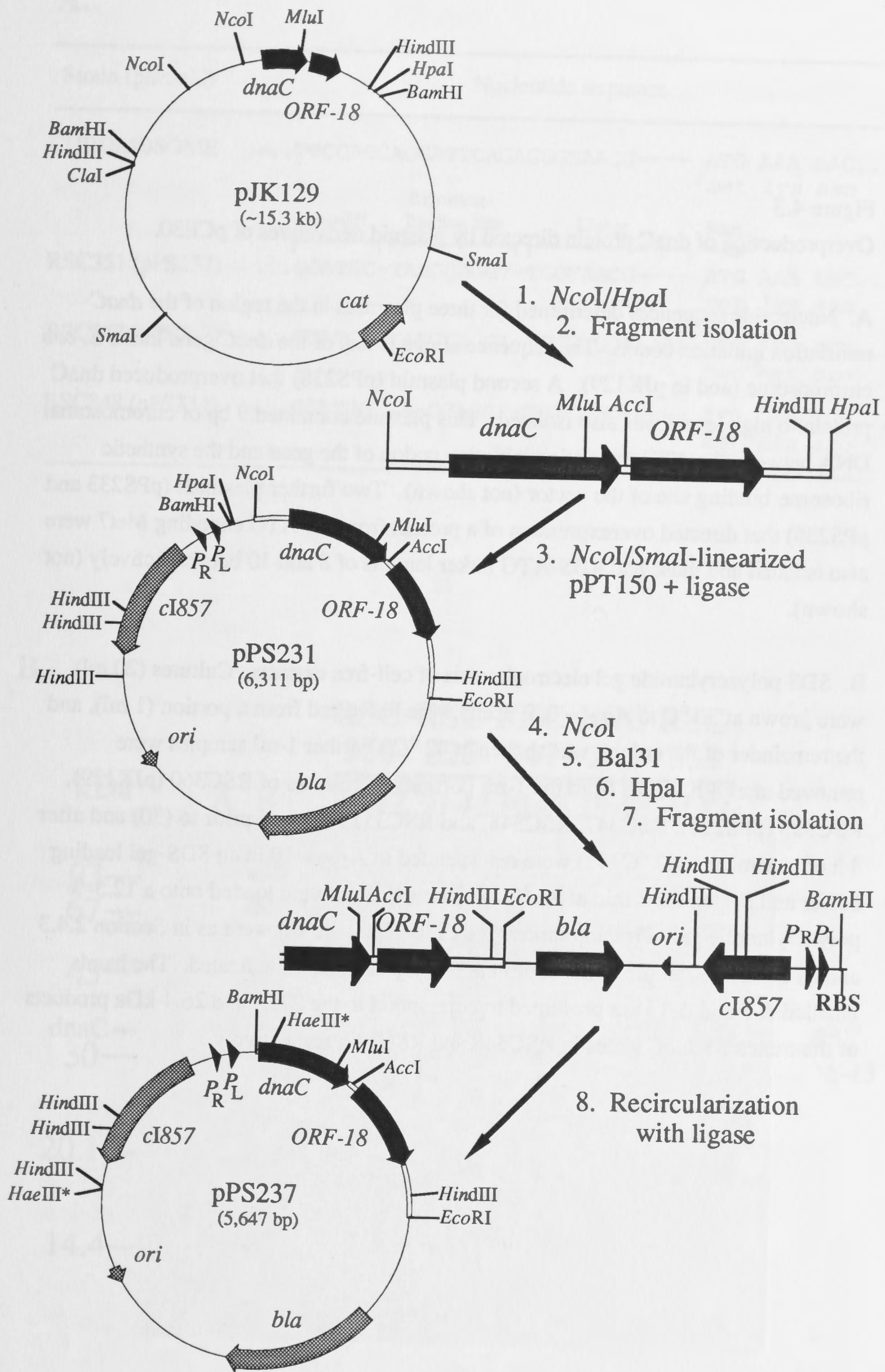


Figure 4.3

Overproduction of *dnaC* protein directed by plasmid derivatives of pCE30.

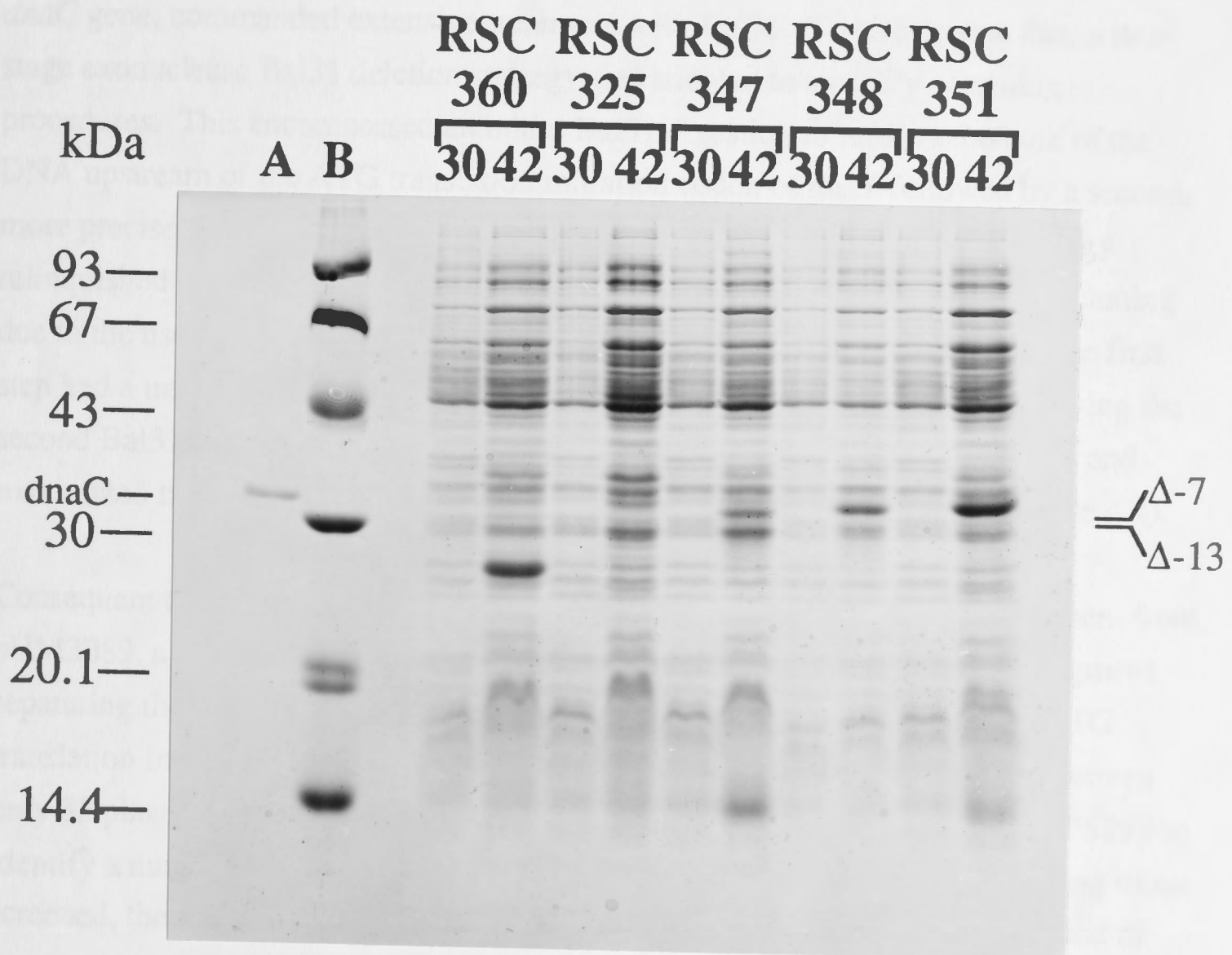
A. Nucleotide sequences determined for three plasmids in the region of the *dnaC* translation initiation codon. The sequence above is that of the *dnaC* gene in the *E. coli* chromosome (and in pJK129). A second plasmid (pPS238) that overproduced *dnaC* protein to high levels was also isolated. This plasmid contained 9 bp of chromosomal DNA between the ATG translation initiation codon of the gene and the synthetic ribosome-binding site of the vector (not shown). Two further plasmids (pPS233 and pPS235) that directed overexpression of a protein from the ATG encoding Met7 were also isolated and these had RBS-ATG linker lengths of 8 and 10 bp, respectively (not shown).

B. SDS polyacrylamide gel electrophoresis of cell-free extracts. Cultures (20 ml) were grown at 30 °C to $A_{595} \sim 0.5$. Cells were harvested from a portion (1 ml), and the remainder of the culture was shaken at 42 °C. Further 1-ml samples were removed after 4 h. Cells from the 1-ml portions of cultures of RSC360 (pJK129), RSC325 (pPS231), RSC347, RSC348, and RSC351 removed prior to (30) and after 4 h of treatment at 42 °C (42) were resuspended to $A_{595} = 10$ in an SDS-gel loading buffer and heated for 2 min at 95 °C. Portions (20 μ l) were loaded onto a 12.5-% polyacrylamide gel. Protein markers (sizes in kDa; lane B.) were as in Section 2.4.3 and an authentic sample of *dnaC* protein ($\sim 4 \mu$ g; lane A.) is indicated. The bands labelled Δ -7 and Δ -13 are presumed to correspond to the 27.2- and 26.4-kDa products of the truncated *dnaC* genes in RSC348 and RSC347, respectively.

A.

Strain (plasmid)	Nucleotide sequence	
CHROMOSOME	...TTCCACCAGGATTCAGAGGGTAACG-----	ATG AAA AAC... ¹ met lys asn
	<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;"> <i>Bam</i>HI ┌───┐ └───┘ </div> <div style="text-align: center;"> Ribosome- Binding Site ┌───┐ └───┘ </div> <div style="text-align: center;"> Linker ┌───┐ └───┘ </div> <div style="text-align: center;"> Start → </div> </div>	
RSC351 (pPS237)	...GGATCC-TAAGGAGGT-TGGTAACG-----	ATG AAA AAC... ¹ met lys asn
RSC347 (pPS232)	...GGATCC-TAAGGAGGT-TGCCTGCAAAAA	ATG ATG CCT... met ¹⁴ met pro
RSC348 (pPS234)	...GGATCC-TAAGGAGGT-TGACCTG-----	ATG CAA CGC... met ⁸ gln arg

B.



although accuracy of such Bal31 treatment can be assured through prior trial experiments, the precision of these digestions are poor. To identify recombinant plasmids with RBS-ATG fusions containing the desired spacing required extensive screening of several hundred transformants. Fortunately, as with primase, high-level expression of the *dnaC* protein induced at restrictive temperatures (for *cl857*) proved lethal to the host. It therefore became a simple task to screen purified transformants by their temperature sensitivity. From this, 96 strains were selected for further analysis by 3' end-labelling studies and restriction endonuclease digests of their plasmid DNA. From these data it was possible to identify those plasmids likely to direct overproduction of the *dnaC* protein at 42 °C (Figure 4.3). Sequencing of these plasmids revealed some that had the new ribosome-binding site fused sufficiently close to separate ATG codons that normally encode Met7 and Met13, that they ought to become new translation initiation codons for NH₂-terminally truncated proteins.

Overproduction of dnaT protein

The nearest convenient restriction endonuclease site upstream of the ATG translation initiation codon of the *dnaT* gene was at a considerable distance (180 bp) and, like the *dnaC* gene, commanded extensive treatment with Bal31. To circumvent this, a two-stage exonuclease Bal31 deletion strategy was adopted to simplify screening procedures. This encompassed an initial Bal31 digestion to remove the bulk of the DNA upstream of the ATG translation initiation codon of *dnaT* followed by a second, more precise, Bal31 deletion on the newly constructed plasmid. Such a strategy relinquished command over orientation of *dnaT*-containing fragments during cloning due to the use of a blunt-ended ligation reaction. The plasmid isolated from the first step had a unique *EcoRI* site downstream of *dnaT*. This, when utilized following the second Bal31 treatment, ensured the correct orientation of the gene in pPS312 and maximized the procurement of potential *dnaT* overexpressing plasmids (Figure 4.4).

Consequent to the Bal31 treatment of the *Clal/HindIII* *dnaT*⁺ containing fragment from pHM3089, a plasmid was isolated with an appropriately-short 50-bp DNA segment separating the *BamHI* site, later utilized in the second Bal31 digest, and the ATG translation initiation codon of *dnaT* (pPS298). From this it was necessary to screen only 32 plasmids produced following Bal31 treatment of *BamHI*-linearized pPS298 to identify a number of excellent *dnaT* overproducing strains (Figure 4.5). Among those screened, there was none that directed overproduction of NH₂-terminal deletions of the *dnaT* protein because Bal31 digestion had not progressed far enough into *dnaT* to

Figure 4.4

Construction of a plasmid, pPS312, that directs overproduction of *dnaT* protein. Details of manipulations are described in Section 4.2.2. The *Bgl*III and *Nco*I sites used in screening of plasmids of anticipated size for the presence and orientation of inserts are centred 248 bp upstream of the *Bam*HI sites in the vector and 256 bp from the 5' end of the *dnaT* coding region, respectively. The *Hpa*II sites (*) used for 3'-end labelling experiments are located 23 bp from the 5' end of the *dnaT* coding region and 1435 bp upstream of the *Bam*HI site in the vector. Although there are 17 other *Hpa*II sites within pPS312, fragments with *Hpa*II cohesive ends are not radioactively end-labelled with 32 P-dATP. The relevant sequences of plasmids selected following these manipulations are given in Figure 4.5.

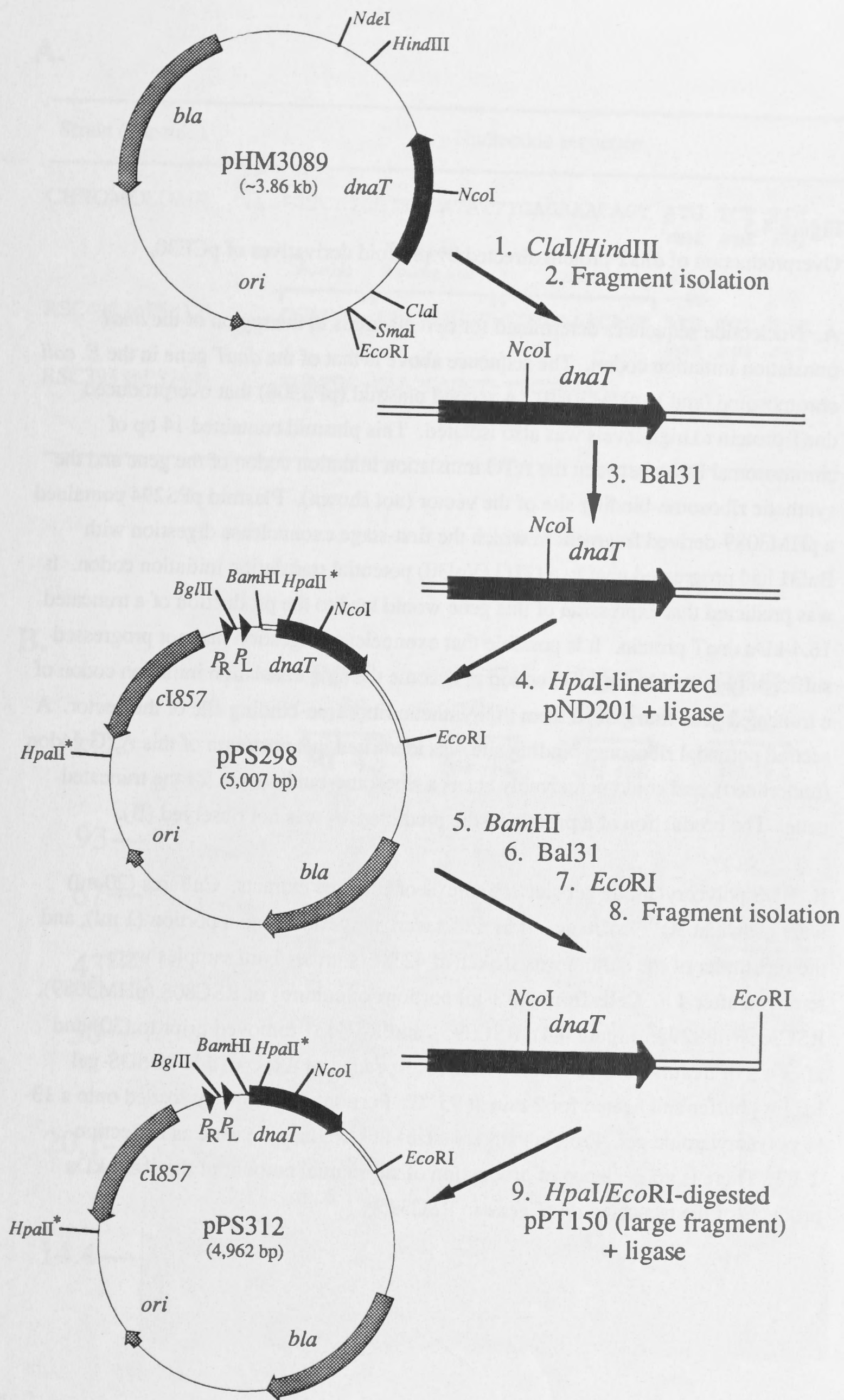


Figure 4.5

Overproduction of *dnaT* protein directed by plasmid derivatives of pCE30.

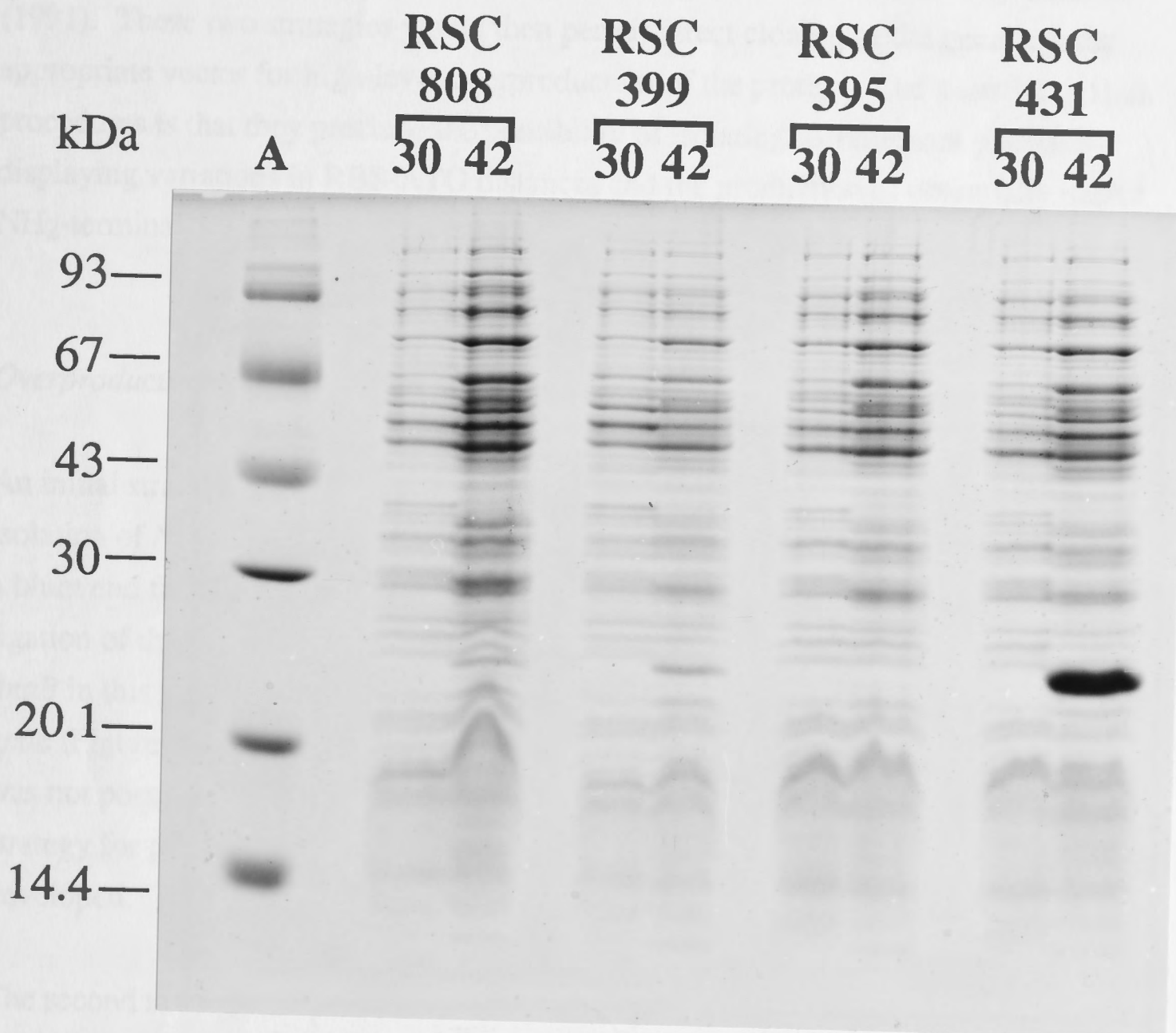
A. Nucleotide sequences determined for two plasmids in the region of the *dnaT* translation initiation codon. The sequence above is that of the *dnaT* gene in the *E. coli* chromosome (and in pHM3089). A second plasmid (pPS308) that overproduced *dnaT* protein to high levels was also isolated. This plasmid contained 14 bp of chromosomal DNA between the ATG translation initiation codon of the gene and the synthetic ribosome-binding site of the vector (not shown). Plasmid pPS294 contained a pHM3089-derived fragment in which the first-stage exonuclease digestion with Bal31 had progressed near to a GTG (Val30) potential translation initiation codon. It was predicted that expression of this gene would lead to the production of a truncated 16.4-kDa *dnaT* protein. It is possible that exonuclease digestion had not progressed sufficiently to enable the ATG codon to become the new translation initiation codon of a truncated gene transcribed from the synthetic ribosome-binding site of the vector. A second potential ribosome-binding site was identified just upstream of this ATG codon (underlined), and could perceivably act as a ribosome-binding site for the truncated gene. The production of a protein of the predicted M_r was not observed (B).

B. SDS polyacrylamide gel electrophoresis of cell-free extracts. Cultures (20 ml) were grown at 30 °C to $A_{595} \sim 0.5$. Cells were harvested from a portion (1 ml), and the remainder of the culture was shaken at 42 °C. Further 1-ml samples were removed after 4 h. Cells from the 1-ml portions of cultures of RSC808 (pHM3089), RSC399 (pPS298; Figure 4.4), RSC395, and RSC431 removed prior to (30) and after 4 h of treatment at 42 °C (42) were resuspended to $A_{595} = 10$ in an SDS-gel loading buffer and heated for 2 min at 95 °C. Portions (20 μ l) were loaded onto a 13% polyacrylamide gel. Protein markers (sizes in kDa; lane A) were as in Section 2.4.3. There is no evidence of production of substantial amount of the 16.4-kDa products of the truncated *dnaT* gene in RSC395.

Strain (plasmid)	Nucleotide sequence
CHROMOSOME	<p>...G TTCATCGTTCCATATTTGAGAAACAGT ATG TCT TCC...</p> <p><i>met ser ser</i></p> <p><i>Bam</i>HI Ribosome- Binding Site Linker Start</p> <p>→</p>
RSC431 (pPS312)	<p>...GGATCC-TAAGGAGGT-TGAGAAACAGT ATG TCT TCC...</p> <p><i>met ser ser</i></p>
RSC395 (pPS294)	<p>...GGATCC-TAAGGAGGT-TTGGCAAAA</p> <p>GCTGAAGGCG GTG TGG TTG...</p> <p><i>val³¹ trp leu</i></p>

B.

RSC	RSC	RSC	RSC
------------	------------	------------	------------



bring the nearest potential translation initiation codon (a GTG encoding Val-31) in proximity to the synthetic ribosome-binding site of the vector.

No plasmid bearing a RBS-ATG linker length <10 bp was isolated in this experiment. The extent of exonuclease digestion during the second Bal31 digest left either >9 ntds upstream of the ATG translation initiation codon of the gene or had progressed into the open reading frame itself. This result probably reflects the propensity of exonuclease Bal31 to quickly degrade AT-rich sequences. This demonstrates a shortcoming of strategies designed to isolate short RBS-ATG linkers by generating Bal31 deletions immediately proximal to the commencement of an open reading frame, as this region of genes is often AT rich. This problem might be overcome by use of ExoIII and mungbean nuclease to generate the deletions. Neither of these exonucleases shows any sequence preference. To eliminate the necessity for extensive screening procedures and bypass difficulties faced using exonuclease digestion, an alternative would be to mutagenize with oligonucleotides that direct the construction of restriction endonuclease sites at the desired position at or upstream of the translation initiation codon of the gene. A similar result could also be achieved using the polymerase chain reaction (PCR) to isolate the coding region of the gene as demonstrated by Godson (1991). These two strategies would then permit direct cloning of the gene into the appropriate vector for high-level overproduction of the protein. The shortfall of such procedures is that they preclude the possibility of isolating recombinant plasmids displaying variations in RBS-ATG distances and the production of potentially useful NH₂-terminal deletions of the protein.

Overproduction of dnaB protein

An initial strategy attempted for overproduction of the dnaB protein required the isolation of *Nde*I-linearized pKA1, Bal31 digestion to remove 44 bp of DNA, creating a blunt end flush with the ATG translation codon of the *dnaB* gene, and blunt-ended ligation of this fragment into the *Sma*I site of pCE30. Successful manipulation of *dnaB* in this fashion would lead to the generation of an *Nco*I site at the start of the gene (Figure 4.6) and provide a perfect tool for further manipulation. Unfortunately it was not possible to clone *dnaB* fragments containing blunt ends and an alternate strategy for generation of a plasmid directing the overproduction of dnaB protein was developed.

The second strategy adopted was similar in many respects to that outlined in the second-phase Bal31-treatment of the *dnaT* gene. By isolating the *dnaB* gene from

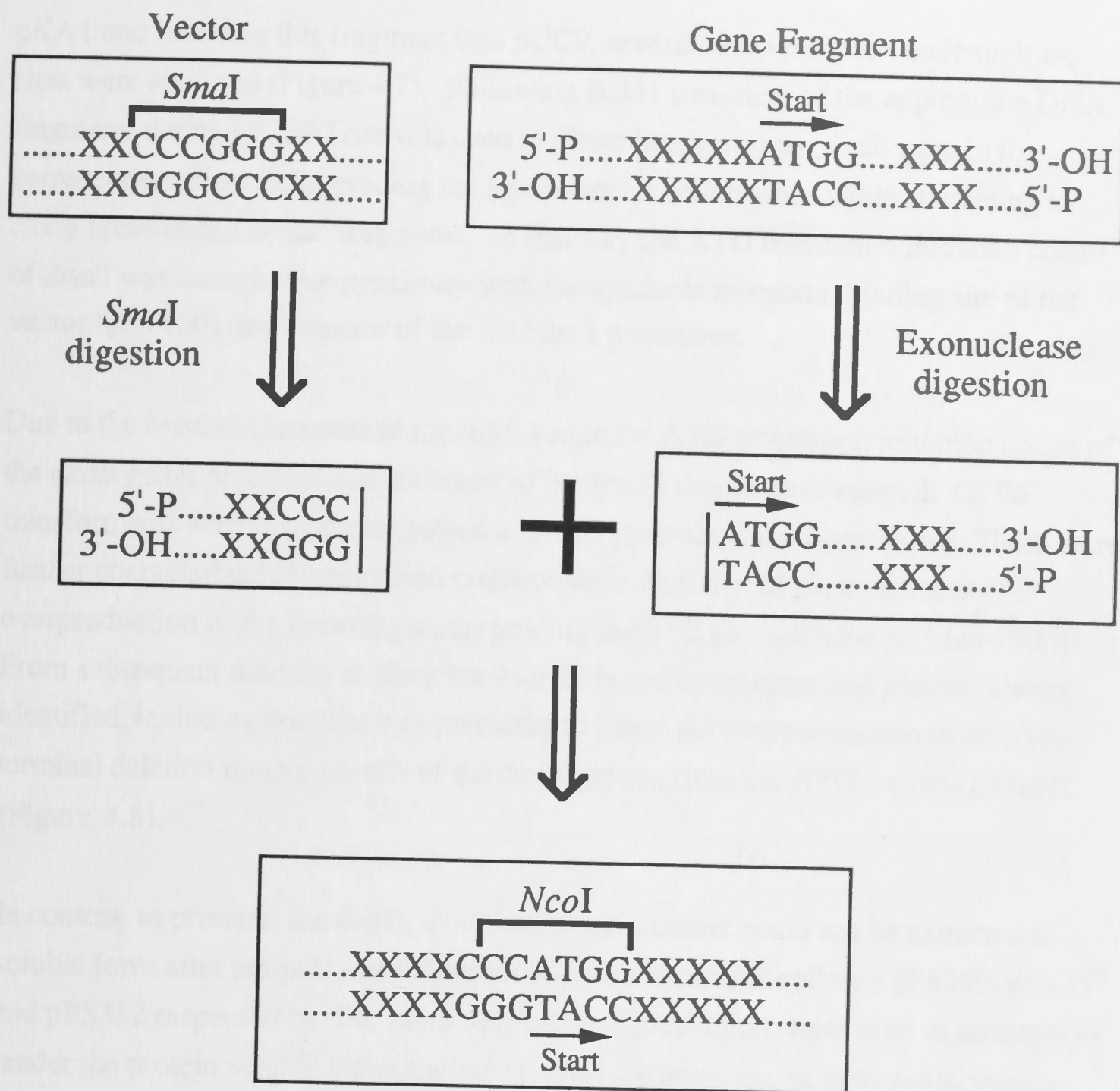


Figure 4.6

The ATG translation initiation codon of the majority of *E.coli* genes allows the chance to utilize this sequence to generate a new restriction endonuclease site at the start of the gene. If the fourth nucleotide of a gene is guanosine, then deleting the DNA upstream of the gene with an exonuclease, such that a blunt end is formed flush with the ATG translation initiation codon, and inserting this fragment into the *Sma*I site of an appropriate vector, allows the formation of a new *Nco*I site. This can then be utilized for further manipulations. The *dnaB* gene, which contains the sequence 5'....ATGG at its 5'-P terminus, was a candidate for such a strategy.

pKA1 and inserting this fragment into pUC9, several new restriction endonuclease sites were acquired (Figure 4.7). Following Bal31 treatment of the appropriate DNA fragment, the new *EcoRI* site was used to direct ligation of the *dnaB* gene in the correct orientation while evading the inefficiencies which arose while attempting to clone blunt-ended *dnaB*⁺ fragments. In this way the ATG translation initiation codon of *dnaB* was brought into proximity with the synthetic ribosome-binding site of the vector (pPT150) downstream of the tandem λ promoters.

Due to the relative closeness of the *NdeI* site to the ATG translation initiation codon of the *dnaB* gene, precision and accuracy of the Bal31 digest were assured. Of 96 transformants screened, eight showed a temperature-sensitive phenotype. These were further characterized by restriction endonuclease digestion of plasmids, and overproduction of the desired protein product at 42 °C as visualized on SDS-PAGE. From subsequent dideoxy sequencing four different overexpression plasmids were identified, including one that was predicted to direct the overproduction of an NH₂-terminal deletion mutant (Δ -42) of the *dnaB* protein from the ATG encoding Met42 (Figure 4.8).

In contrast to primase, the *dnaB*, *dnaC* and *dnaT* proteins could not be extracted in soluble form after temperature induction of *E. coli* strains containing pPS359, pPS237 and pPS312 respectively. For *dnaB* several lysis procedures were used in attempts to render the protein soluble including mechanical shearing, use of detergents, freeze-thawing and lysozyme lysis under varying conditions of pH, ionic strength and temperature with little or no success. Altering the length of time of the induction at 42 °C and changing the temperature at which induction was performed between 41 °C and 45 °C also had no effect on the solubility of the protein. Extraction of the *dnaC* and *dnaT* protein was examined less extensively, but the results were similarly discouraging (data not shown).

The *dnaB* protein was partially solubilized following disruption of thermally-induced cells containing pPS359 in a French press at >18,000 psi. This protein was then partially purified in trial preparations by precipitation with solid ammonium sulfate and DEAE-cellulose anion-exchange chromatography and was partially active in replication assays (results not shown). The *dnaB* protein from these preparations was present in several uncharacterized forms, as evidenced by its very broad elution on a Sephacryl S-400 gel filtration column, was badly contaminated with DNA and was difficult to separate from other contaminating proteins. The material isolated in this fashion was judged unsuitable for use in structure/function experiments or replication assays.

Figure 4.7

Construction of a plasmid, pPS353, that directs overproduction of *dnaB* protein.

Details of manipulations are described in Section 4.2.3. The *Nco*I site used in screening of plasmids of anticipated size for the presence of inserts, and approximate end-points of *Bal*31 digestion, is centred 901 bp from the 5' end of the *dnaB* gene. The plasmid pPS359 (6,110 bp) was constructed from pPS353 by deletion of the small *Sma*I fragment. The relevant sequences of plasmids selected following these manipulations are given in Figure 4.8.

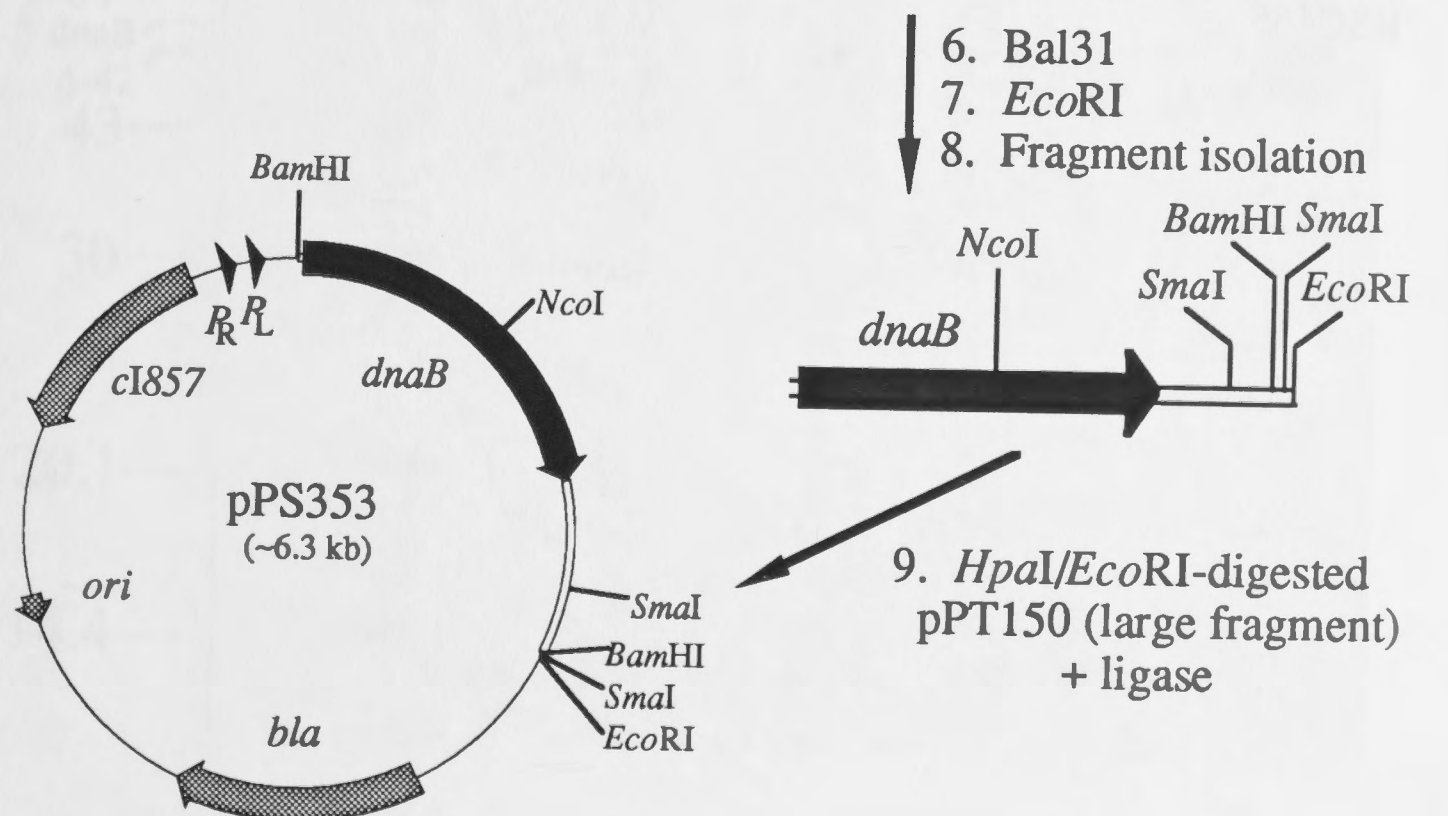
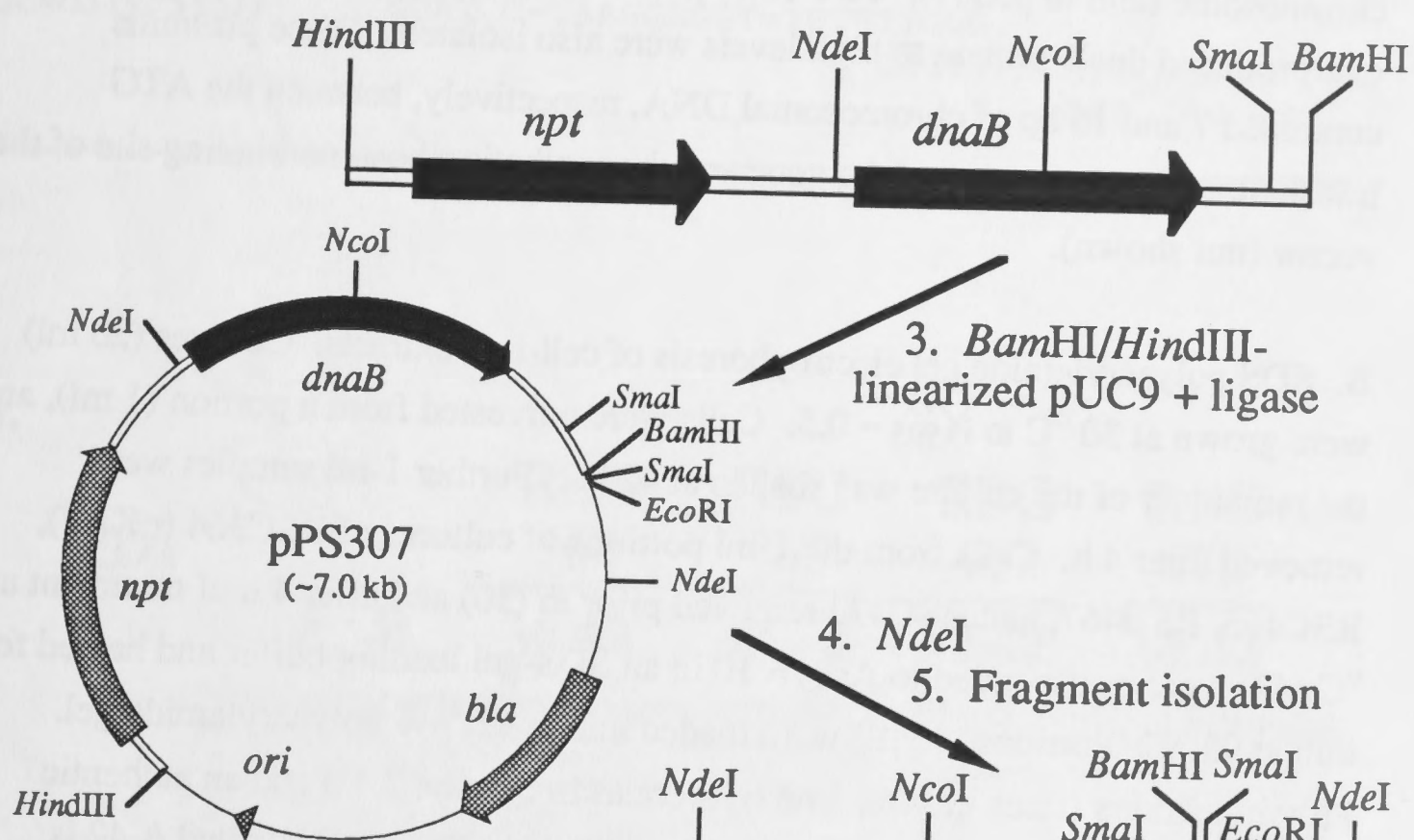
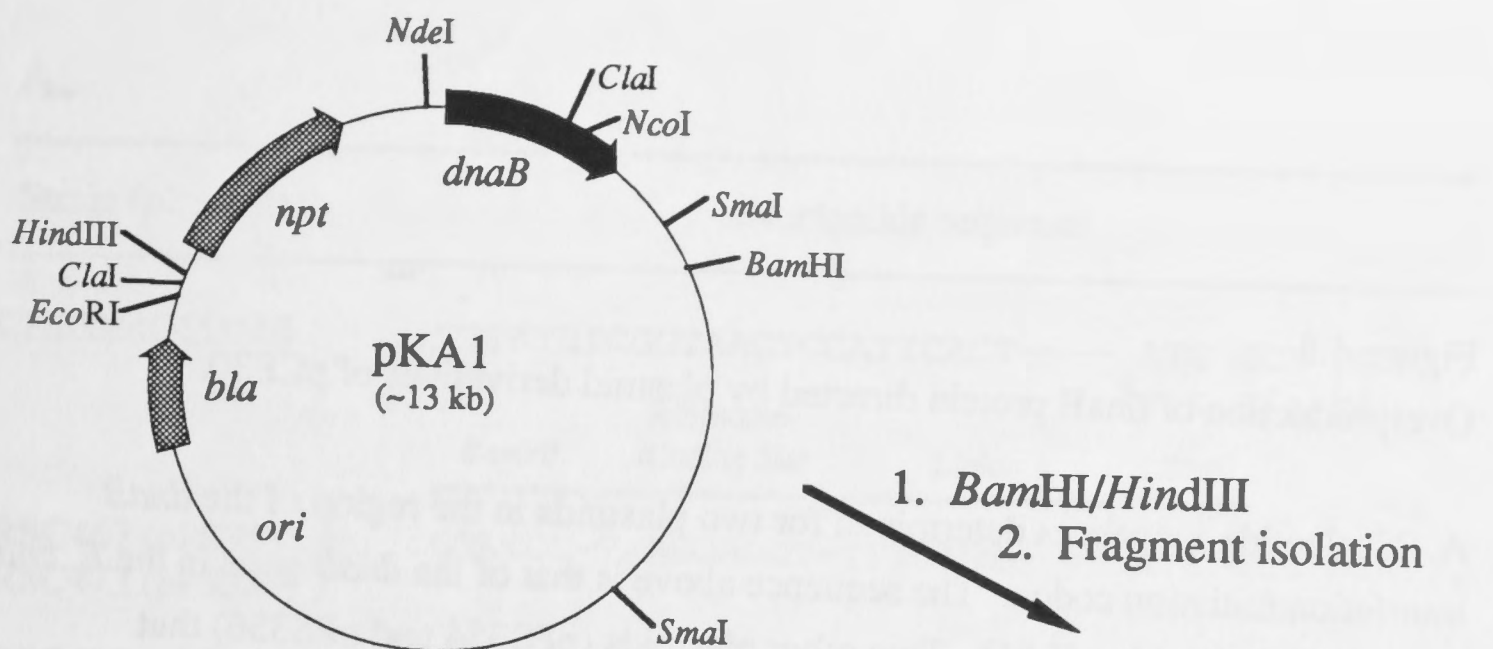


Figure 4.8

Overproduction of *dnaB* protein directed by plasmid derivatives of pCE30.

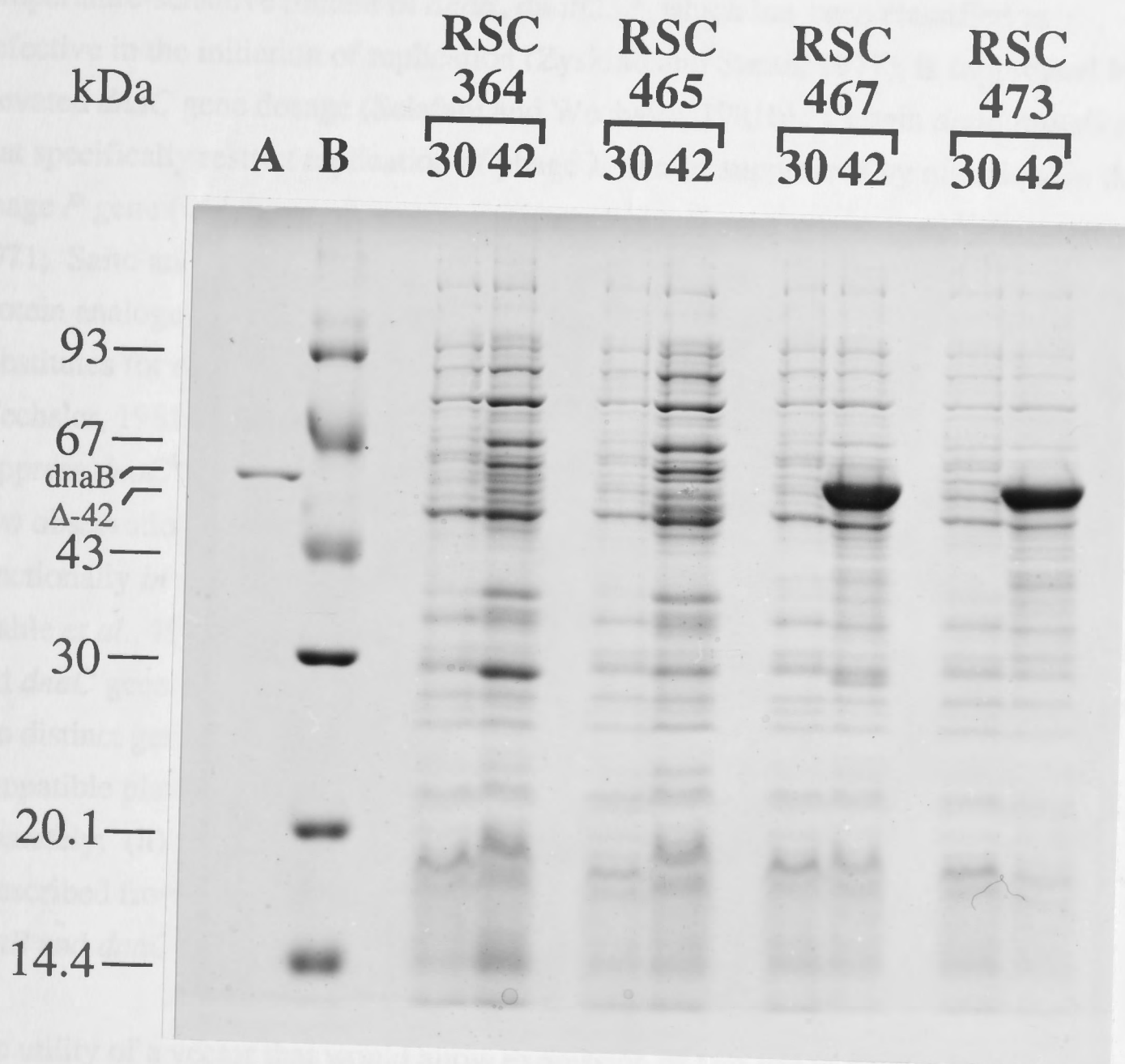
A. Nucleotide sequences determined for two plasmids in the region of the *dnaB* translation initiation codon. The sequence above is that of the *dnaB* gene in the *E. coli* chromosome (and in pKA1). Two other plasmids (pPS354 and pPS356) that overproduced *dnaB* protein to high levels were also isolated. These plasmids contained 7 and 16 bp of chromosomal DNA, respectively, between the ATG translation initiation codon of the gene and the synthetic ribosome-binding site of the vector (not shown).

B. SDS polyacrylamide gel electrophoresis of cell-free extracts. Cultures (20 ml) were grown at 30 °C to $A_{595} \sim 0.5$. Cells were harvested from a portion (1 ml), and the remainder of the culture was shaken at 42 °C. Further 1-ml samples were removed after 4 h. Cells from the 1-ml portions of cultures of RSC364 (pKA1), RSC465, RSC467, and RSC473 removed prior to (30) and after 4 h of treatment at 42 °C (42) were resuspended to $A_{595} = 10$ in an SDS-gel loading buffer and heated for 2 min at 95 °C. Portions (20 μ l) were loaded onto a 12.5-% polyacrylamide gel. Protein markers (sizes in kDa; lane B) were as in Section 2.4.3 and an authentic sample of *dnaB* protein (~ 4 μ g; lane A) is indicated. The band labelled Δ -42 is presumed to correspond to the 47.9-kDa product of the truncated *dnaB* gene in RSC465.

A.

Strain (plasmid)	Nucleotide sequence
CHROMOSOME	...TTTATCTCGGTAACCTCCATTCACT----ATG GCA GGA... met ala gly
	<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;">BamHI</div> <div style="text-align: center;">Ribosome- Binding Site</div> <div style="text-align: center;">Linker</div> <div style="text-align: center;">Start →</div> </div>
RSC467 (pPS353) RSC473 (pPS359)	...GGATCC-TAAGGAGGT-TTTCACCT----ATG GCA GGA... met ala gly
RSC465 (pPS351)	...GGATCC-TAAGGAGGT-TGTGTTGGG CGGTTTA ATG CTA GAT... met ⁴³ leu asp

B.



Co-overproduction of dnaB and dnaC protein

The insolubility of the overproduced protein that accumulates following thermal induction in these strains is likely the result of the overproduced protein being unable to adopt a native structure under the conditions in which it is expressed. The only alternative to extracting the protein from these insoluble aggregates would therefore involve directing the formation of a soluble conformation from the nascent polypeptide during production of the protein. Two avenues that may promote production of the desired protein conformation from nascent amino acid chains include lowering the temperature of induction to physiological levels and expressing other proteins which possibly interact with it *in vivo*. These processes rely on the assumption that either the protein progresses to an energetically favourable but insoluble conformation at elevated temperatures or is otherwise stabilized by the presence of closely associated polypeptides. Both proposals were investigated and involved the development of novel recombinant plasmids.

Genetic evidence suggests that dnaB protein and dnaC protein interact *in vivo*. A temperature-sensitive mutant of *dnaB*, *dnaB252*, which has been classified as defective in the initiation of replication (Zyskind and Smith, 1977), is suppressed by elevated *dnaC* gene dosage (Sclafani and Wechsler, 1981b). Certain *dnaB* mutations that specifically restrict replication of phage λ can be suppressed by mutations in the phage *P* gene (which encodes a dnaC-like protein; Georgopoulos and Herskowitz, 1971; Saito and Uchida, 1977; Wickner, 1979; Klein *et al.*, 1980). The dnaB protein analogue encoded by the *ban* gene of bacteriophage P1, which completely substitutes for dnaB function *in vivo*, can suppress several *dnaC* alleles (Sclafani and Wechsler, 1981a). Furthermore, several mutations in *dnaB* have been isolated that suppress *dnaC*^{ts} mutations in *Salmonella* (Maurer *et al.*, 1984). Consistent with *in vivo* observations, purified *E. coli* dnaB and dnaC proteins interact physically and functionally *in vitro* to form a tight, isolable complex (Kobori and Kornberg, 1982c; Wahle *et al.*, 1989a). With this knowledge, the simultaneous expression of the *dnaB* and *dnaC* genes was investigated. Three methods of conjointly overexpressing these two distinct genes in a single cell could be employed: (i) maintaining in the cell two compatible plasmids directing the overproduction of dnaB and dnaC proteins separately; (ii) constructing a plasmid with the *dnaB* and *dnaC* genes separately transcribed from independent promoters; and (iii) engineering a plasmid where the *dnaB* and *dnaC* genes are co-transcribed from the same promoter.

The utility of a vector that would allow expression of two genes simultaneously led to the development of pPS436, a new vector derivative of pCE30 that enables the

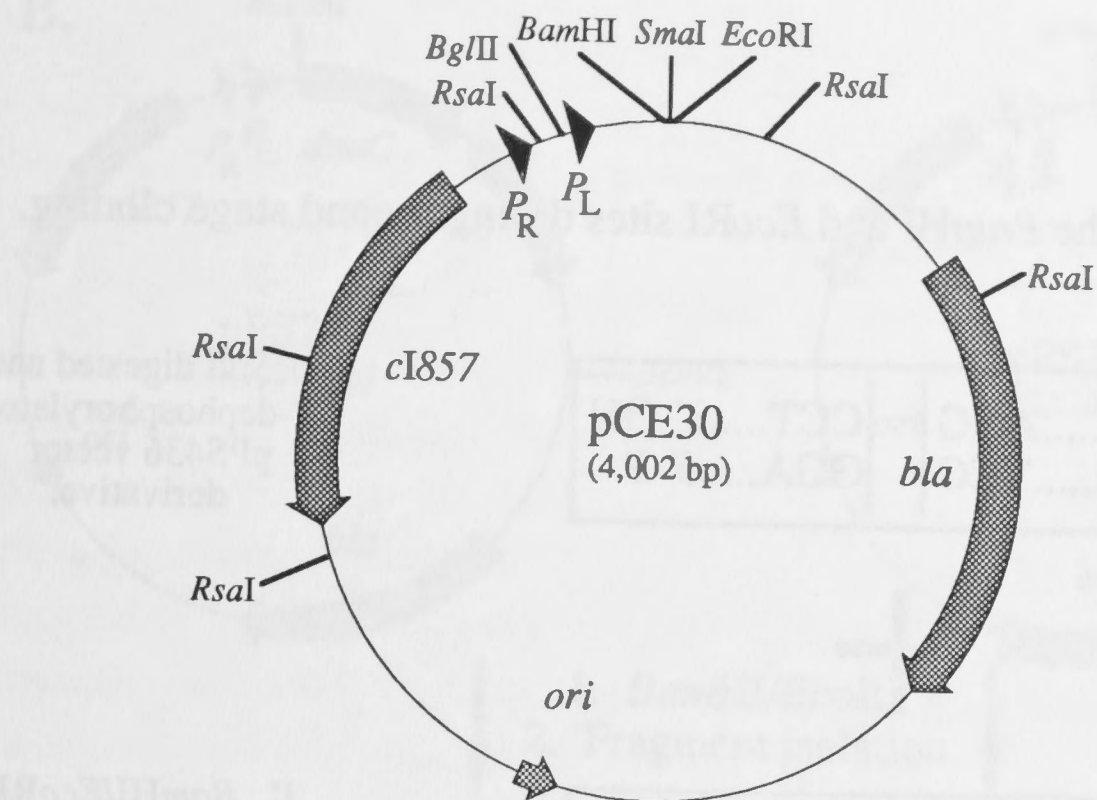
insertion of genes downstream of either P_R or P_L . The strategy adopted for the construction of pPS436 involved insertion of a polylinker sequence near the unique *Bgl*III site that lay between the P_R or P_L promoters in pCE30 (Figure 4.9). It would have been possible to simply insert the desired polylinker into the *Bgl*III site, however, there were concerns that a strong ribosome-binding site derived from the *cII* gene upstream of *Bgl*III may interfere with the level of gene expression from a gene inserted into a polylinker at the *Bgl*III site. The proximity of two strong ribosome-binding sites often leads to suppression of translation initiation from one, presumably by size exclusion of the ribosome (N.E. Dixon, unpublished). To eliminate this possibility a strategy was developed that made use of the *Rsa*I site between P_R and the *cII* ribosome-binding site in conjunction with the *Bgl*III site of pCE30 to delete the *cII* ribosome-binding site and provide a site for insertion of the polylinker. The polylinker sequence was derived from the multiple cloning sites of pMTL20P (Chambers *et al.*, 1988; Figure 2.1) and was chosen for its extensive series of restriction sites which were largely unique relative to pCE30. This region served as a perfect locale for the ligation of genes under the direct control of λP_R .

A consequence of molecular cloning using the pCE30 derivatives is the manufacture of synthetic RBS-gene fusions within a *Bam*HI/*Eco*RI fragment. This feature simplifies any further plasmid manipulations that may be required. Isolating the *dnaB* and *dnaC* genes from the novel expression vectors pPS359 and pPS237, respectively, and their insertion downstream of either λ promoter in pPS436 was therefore a relatively simple task. This process was also assisted by the reformation of both the *Bam*HI and *Eco*RI sites following the ligation of the blunt-ended *Bam*HI/*Eco*RI RBS-gene fusion cassettes from pPS359 and pPS237 into the *Stu*I site downstream of λP_R (Figure 4.10).

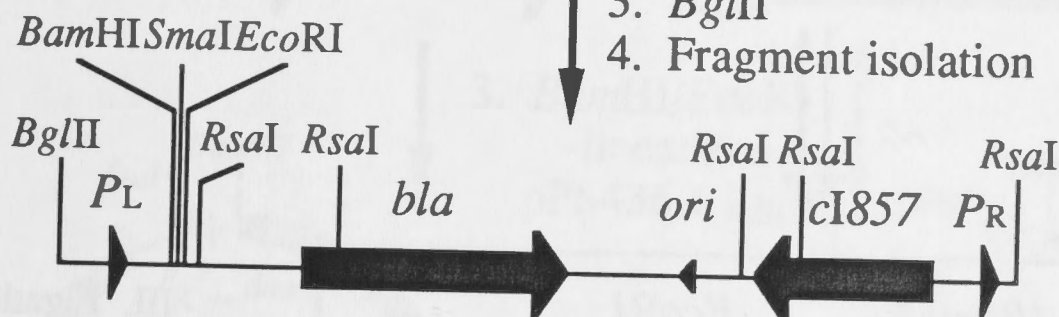
Of the two plasmids isolated from this exercise only pPS554 overproduced both *dnaB* and *dnaC* proteins concurrently (Figure 4.13). All of 5 plasmid isolates maintaining the *dnaC* gene under λP_R and the *dnaB* gene under λP_L failed to direct over-production of the *dnaC* gene product. Considering that under P_L the expression of the *dnaC* gene seemed quite efficient and that under λP_R the *dnaB* gene is expressed equally as efficiently, the likely reason for the inability to over-produce *dnaC* protein under the direction of λP_R is a translational problem. Sequences immediately downstream of λP_R and/or upstream of λP_L may induce the development of secondary structural elements with the *dnaC* gene on the transcript and interfere with the normal translation of *dnaC* from the mRNA. As expected, the *dnaB* protein produced following induction of a strain bearing this plasmid that failed to simultaneously over-produce *dnaC* was totally insoluble.

Figure 4.9

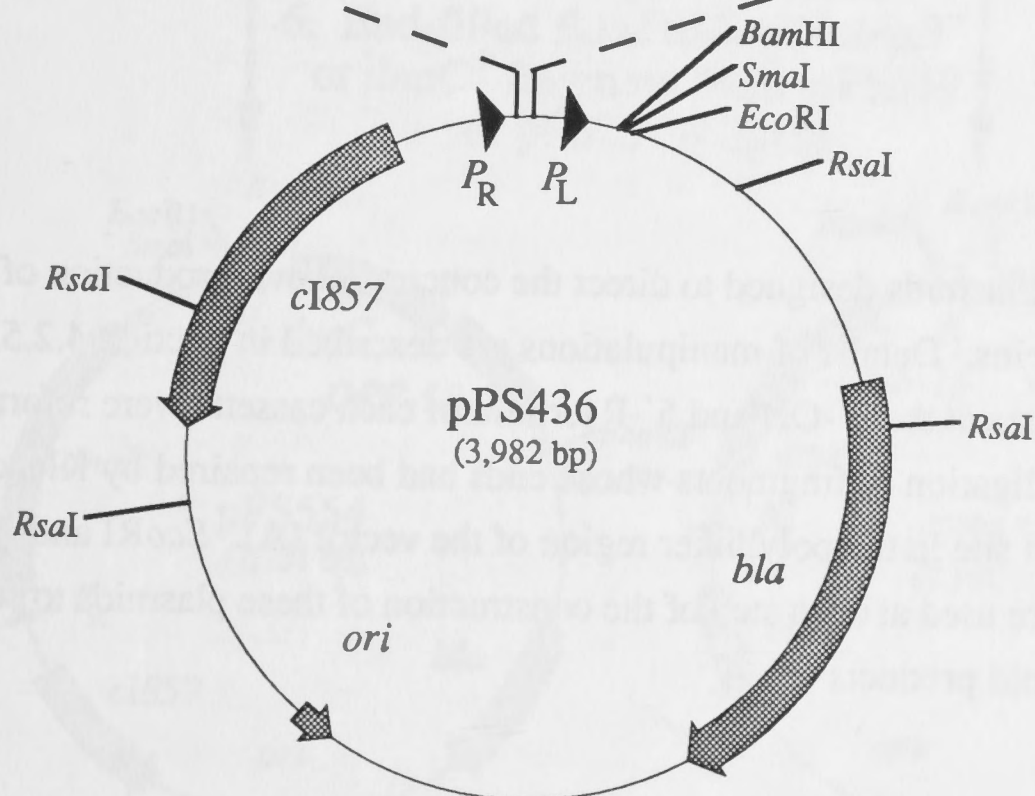
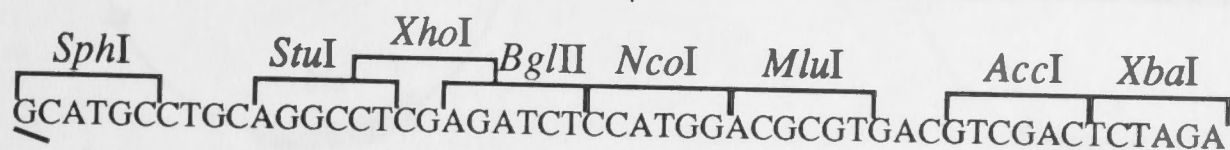
Construction and genetic and physical map of pPS436 (3982 bp), a derivative of pCE30 (Section 2.2.1), showing the locations of the two λ promoters P_R and P_L , the *bla* and λ *ci857* genes, and the restriction sites available for cloning. The extensive polylinker region separating P_R and P_L was obtained from pMTL20P (Section 2.2.1) and introduced as described (Section 4.2.5.1). The presence of restriction sites was confirmed by digestion with the appropriate restriction endonucleases. Orientation of the polylinker was confirmed by sequencing the relevant section of a plasmid derivative (PC-3) with a deletion between the *Eco*RI and *Stu*I sites.



1. *RsaI*-linearized (by partial digest)
2. Fragment isolation
3. *BglII*
4. Fragment isolation



5. *HindIII*-end-filled/*BamHI* 58-bp polylinker from pMTL20P + ligase



A.

Reformation of the *Bam*HI and *Eco*RI sites during second stage cloning.

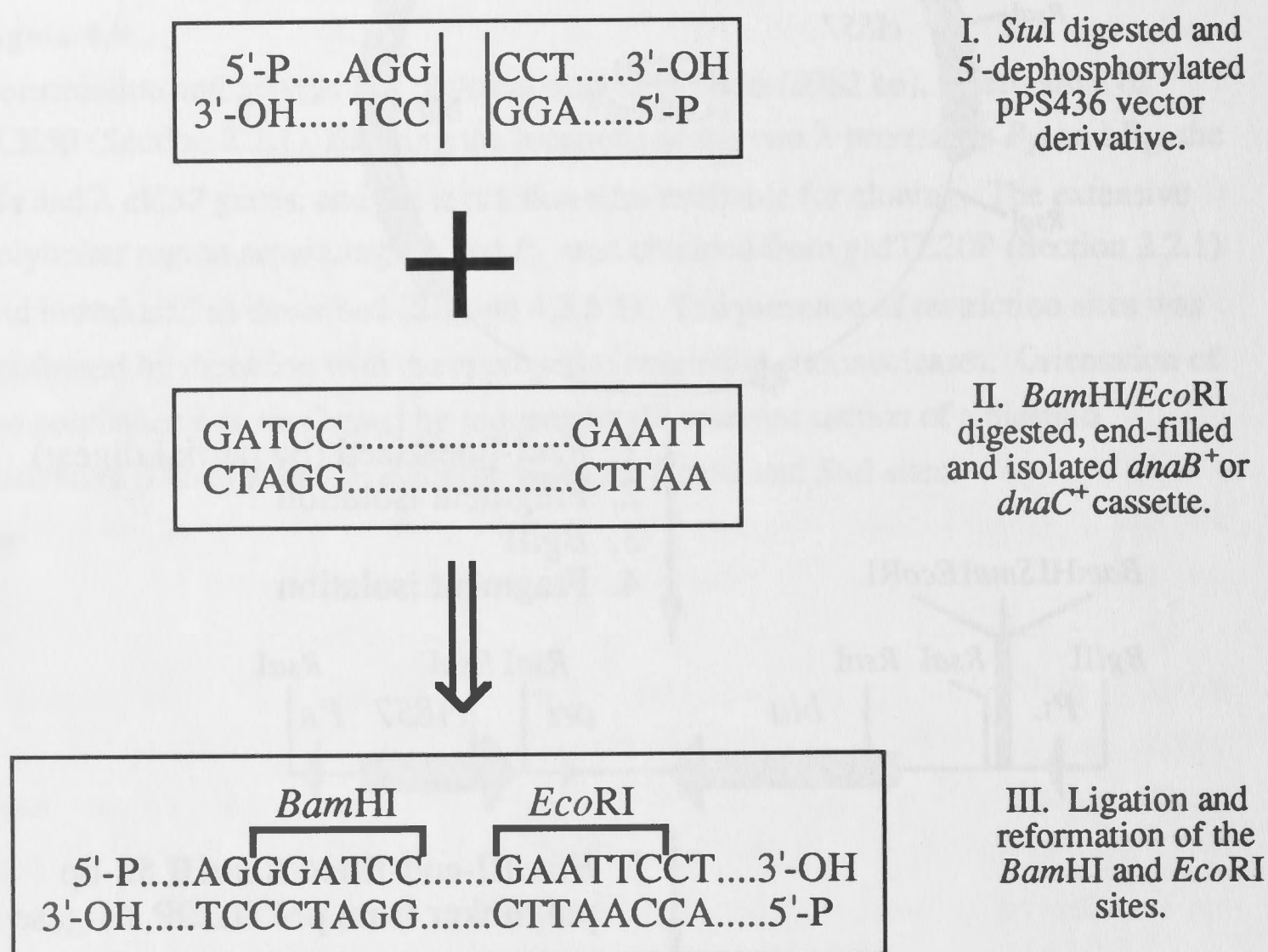
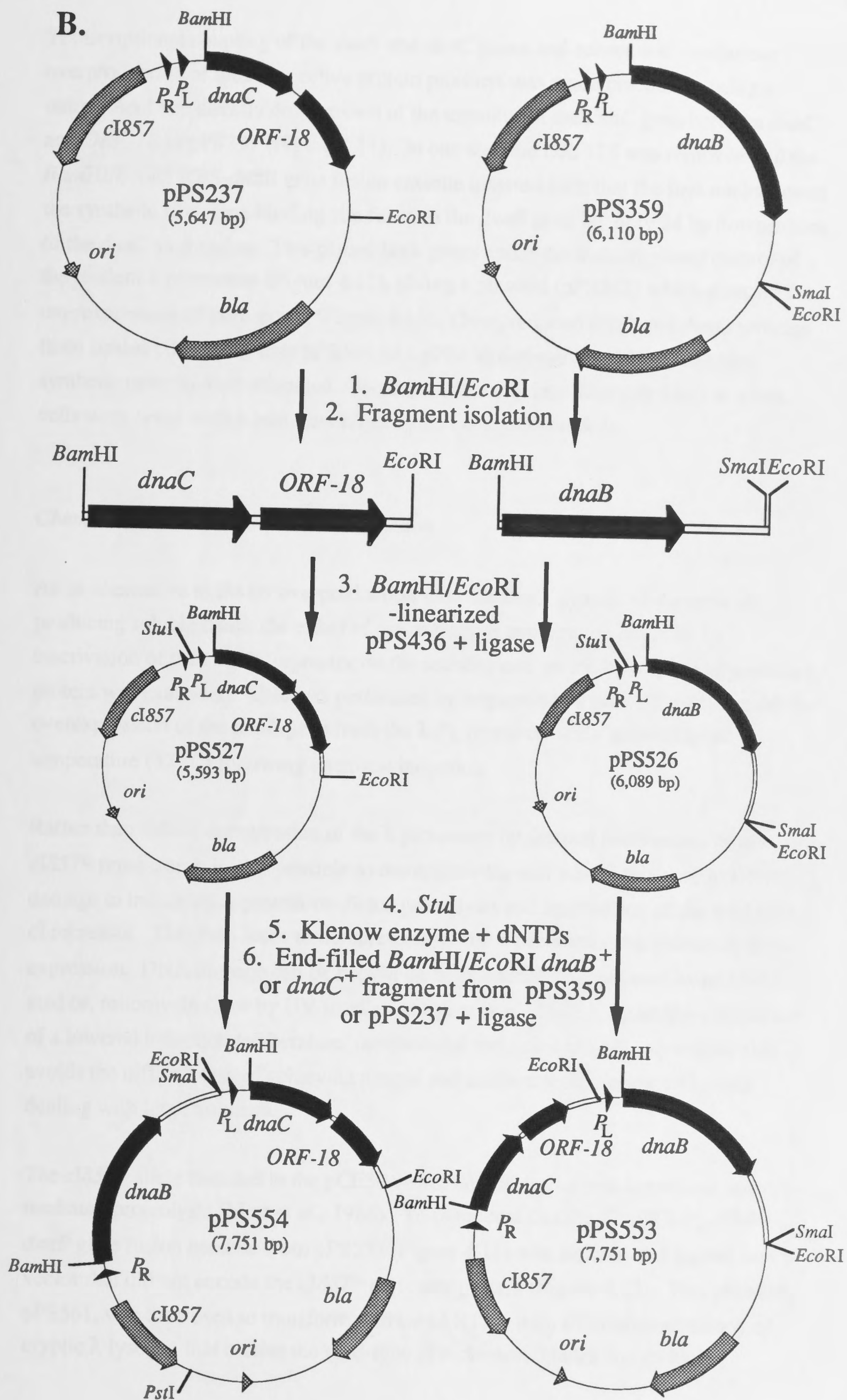


Figure 4.10

Construction of two plasmids designed to direct the concurrent overproduction of both *dnaB* and *dnaC* proteins. Details of manipulations are described in Section 4.2.5.2. The *Bam*HI and *Eco*RI sites at the 3'-OH and 5'-P termini of each cassette were reformed following blunt-end ligation of fragments whose ends had been repaired by Klenow enzyme, into the *Stu*I site in the polylinker region of the vector (A). *Eco*RI and *Bam*HI digests were therefore used at each step of the construction of these plasmids to screen for the desired plasmid products (B).

B.



Transcriptional coupling of the *dnaB* and *dnaC* genes and subsequent concurrent overproduction of their respective protein products was made practicable using a unique *AccI* site directly downstream of the terminus of the *dnaC* gene between *dnaC* and *ORF-18* in pPS237 (Figure 4.11). In one step the *ORF-18* was removed and the *Bam*HI/*Eco*RI RBS-*dnaB* gene fusion cassette inserted such that the first nucleotide of the synthetic ribosome-binding site fused to the *dnaB* gene lay only 24 bp downstream of the *dnaC* stop codon. This placed both genes under the transcriptional control of the tandem λ promoters (Figure 4.11), giving a plasmid (pPS562) which directed the overexpression of both genes (Figure 4.13). Overproduced *dnaB* and *dnaC* proteins from strains containing both pPS554 (the pPS436 derivative) and pPS562 (the synthetic operon) were extracted almost quantitatively into a soluble fraction when cells were lysed with a heat-shock/lysozyme lysis (Section 5.3).

Chemical induction of dnaB overproduction

As an alternative to the co-overproduction with the *dnaC* protein, as a means of producing soluble *dnaB*, the effect of the restrictive temperature required for inactivation of the *cI857^{ts}* repressor on the stability and solubility of the overproduced protein was examined. This was performed by engineering a plasmid that directed the overexpression of the *dnaB* gene from the λP_L promoter at the physiological temperature (37 °C) following chemical induction.

Rather than induce derepression of the λ promoters by thermal inactivation of the *cI857^{ts}* repressor, it is also possible to manipulate the cell's SOS response to DNA damage to induce *recA* protein-mediated proteolysis and inactivation of the wild-type *cI* repressor. This then leads to derepression of the λ promoters and ultimately gene expression. DNA damage can be brought on by chemical reagents such as nalidixic acid or, mitomycin C, or by UV irradiation (Mott *et al.*, 1985). Apart from the virtue of a lowered induction temperature, the chemical induction of gene expression also avoids the difficult task of achieving a rapid and uniform temperature shift when dealing with large volumes.

The *cI857^{ts}* allele encoded in the pCE30 vector and its derivatives is resistant to *recA*-mediated proteolysis (Mott *et al.*, 1985). To overcome this the *Eco*RI λP_L -RBS-*dnaB* gene fusion cassette from pPS553 (Figure 4.10) was isolated and ligated into a vector that did not encode the *cI857^{ts}* repressor protein (Figure 4.12). This plasmid, pPS561, was then used to transform a strain (AR120) with a chromosome copy of cryptic λ lysogen that carries the wild-type *cI* repressor. Under non-inducing

Figure 4.11

Construction of a plasmid, pPS562, directing co-transcription of *dnaB* and *dnaC* in a synthetic operon. Co-transcription of the *dnaB* and *dnaC* genes was made possible by ligating *dnaB* immediately behind *dnaC* under the direction of P_R-P_L . The resultant plasmid construct was predicted to contain a short 23-bp linker separating the ochre stop codon of the *dnaC* gene and the synthetic ribosome-binding site fused upstream of *dnaB*. The 5'-P and 3'-OH termini of the coding regions of *dnaB* and *dnaC*, respectively, are indicated by arrows above the sequence. The TAA ochre stop codon of *dnaC* and ATG translation initiation codon of the *dnaB* gene are indicated by shaded box regions.

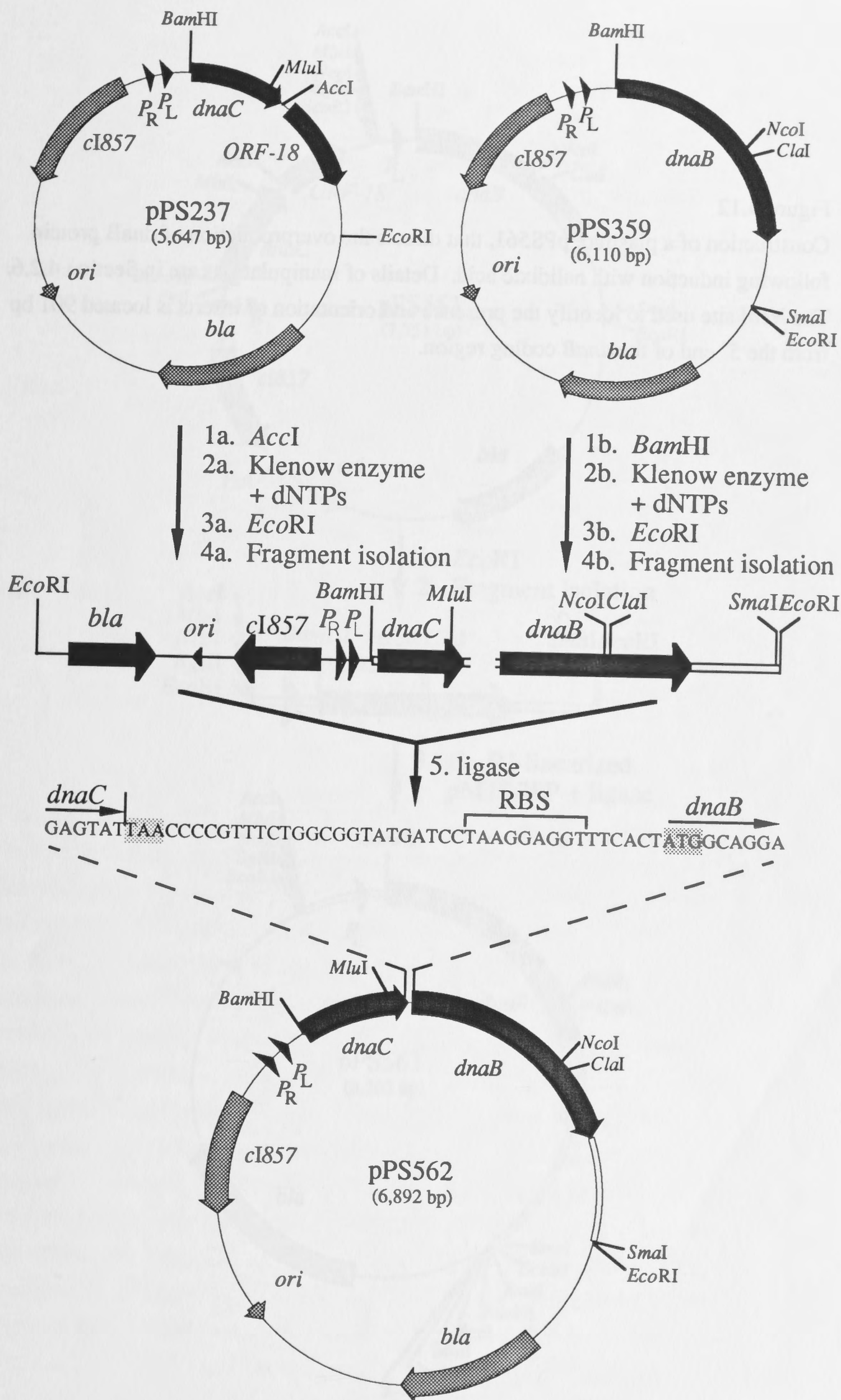
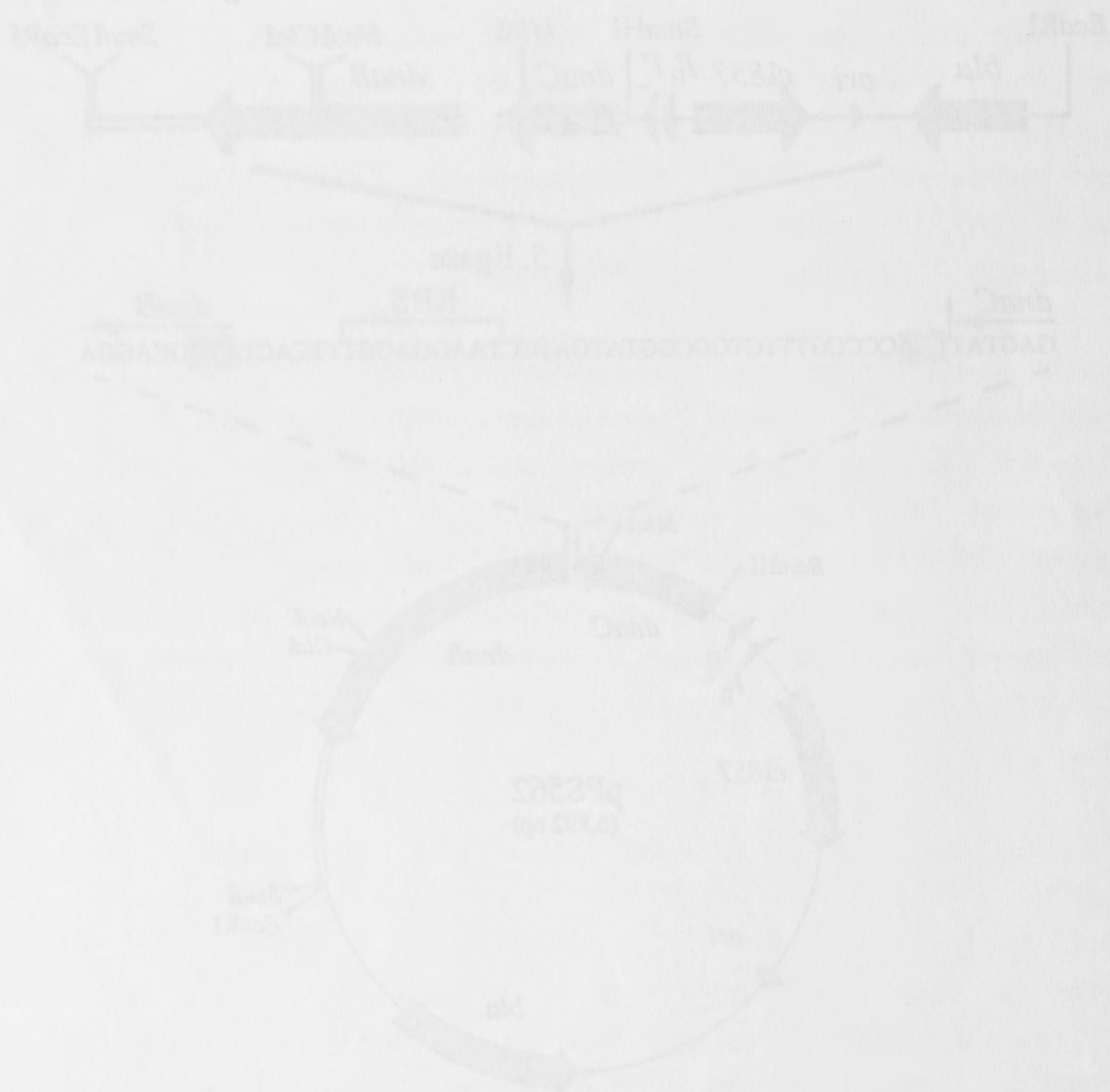
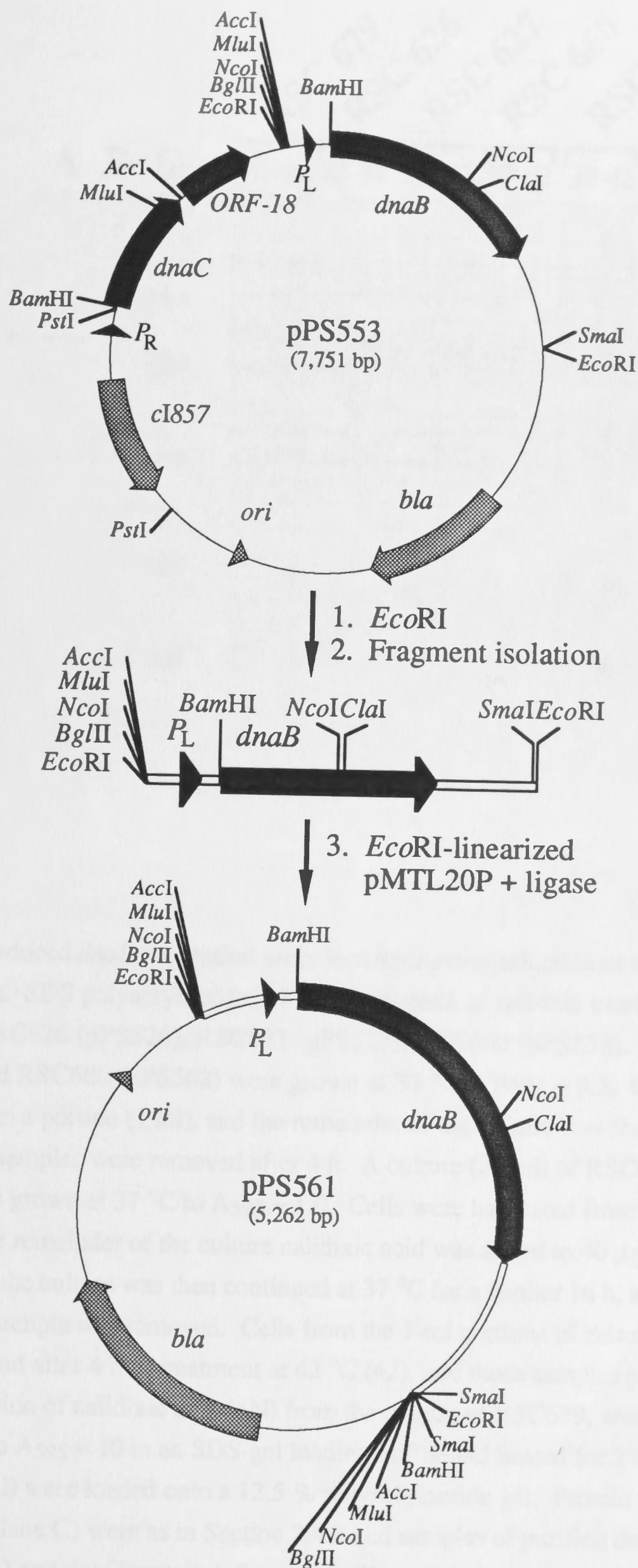


Figure 4.12

Construction of a plasmid, pPS561, that directs the overproduction of *dnaB* protein following induction with nalidixic acid. Details of manipulations are in Section 4.2.6. The *Nco*I site used to identify the presence and orientation of inserts is located 901 bp from the 5' end of the *dnaB* coding region.





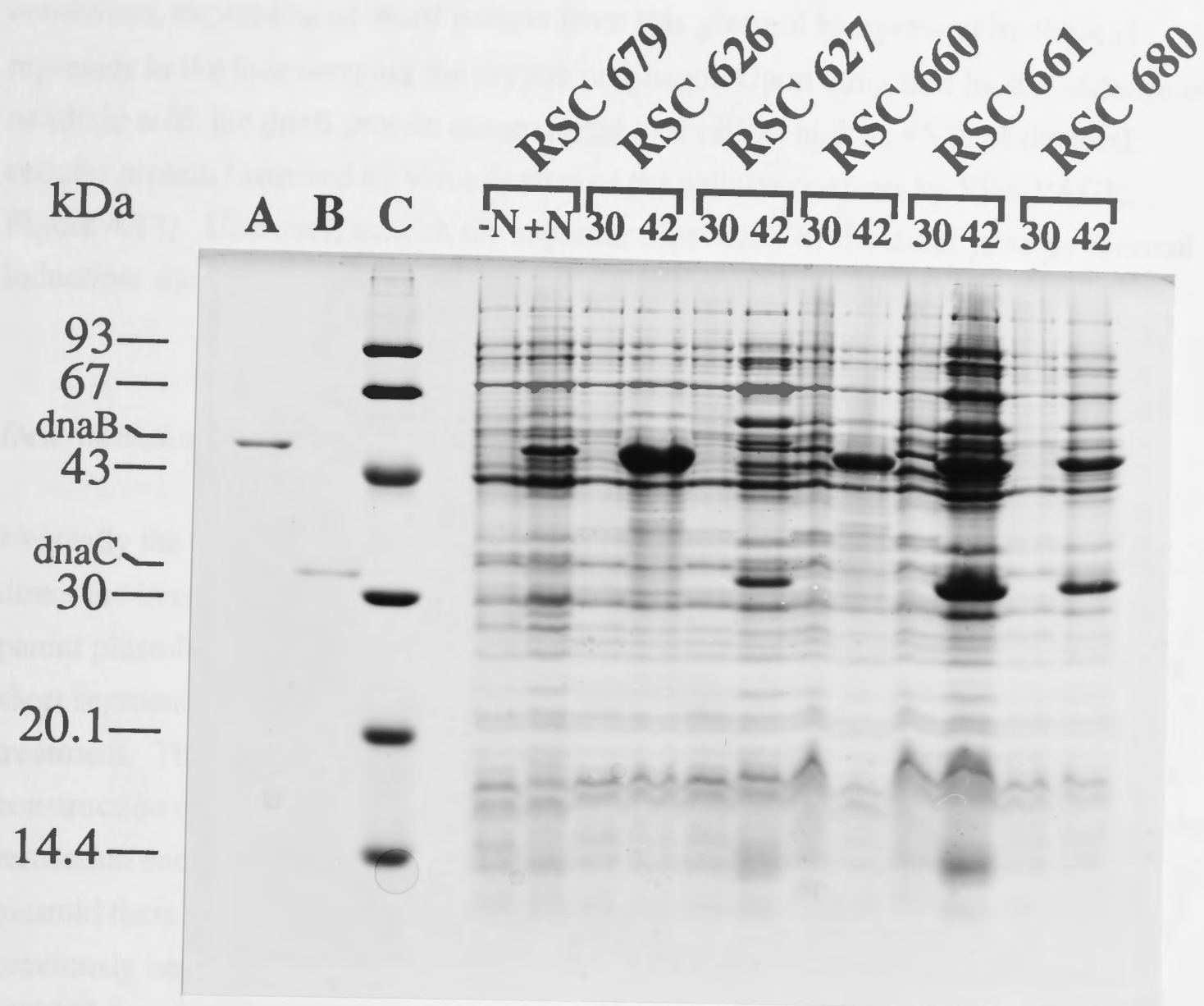


Figure 4.13

Chemically induced *dnaB* expression and concurrent overproduction of the *dnaB* and *dnaC* proteins. SDS polyacrylamide gel electrophoresis of cell-free extracts. Cultures (20 ml) of RSC626 (pPS526), RSC627 (pPS527), RSC660 (pPS553), RSC661 (pPS554), and RSC680 (pPS562) were grown at 30 °C to $A_{595} \sim 0.5$. Cells were harvested from a portion (1 ml), and the remainder of the culture was shaken at 42 °C. Further 1-ml samples were removed after 4 h. A culture (20 ml) of RSC679 (pPS561) was grown at 37 °C to $A_{595} \sim 1.0$. Cells were harvested from a portion (1 ml), and to the remainder of the culture nalidixic acid was added to 40 µg/ml. Incubation of the culture was then continued at 37 °C for a further 16 h, at which time another 1-ml sample was removed. Cells from the 1-ml portions of cultures removed prior to (30) and after 4 h of treatment at 42 °C (42), and those samples prior to (-N) and after addition of nalidixic acid (+N) from the culture of RSC679, were resuspended to $A_{595} = 10$ in an SDS-gel loading buffer and heated for 2 min at 95 °C. Portions (20 µl) were loaded onto a 12.5-% polyacrylamide gel. Protein markers (sizes in kDa; lane C) were as in Section 2.4.3 and samples of purified *dnaB* protein (~4 µg; lane A) and *dnaC* protein (~2 µg; lane B) are indicated.

conditions, expression of *dnaB* protein from this plasmid is repressed by the λ *cI* repressor in the host carrying the cryptic prophage. Upon induction by the addition of nalidixic acid, the *dnaB* protein accumulated to levels as high as ~5 % of the total cellular protein (assessed by visualization of the cellular contents by SDS-PAGE; Figure 4.13). However, as with the separate expression of the *dnaB* gene by thermal induction, the protein produced in this manner was also completely insoluble.

Determination of the complete sequence of pPS359

Normally the complete nucleotide sequence of recombinant derivatives of pCE30 that direct the overexpression of inserted genes can be deduced from the sequence of the parent plasmid vector, the reported sequence of the gene, and the determination of the short segments of DNA remaining at either end of the inserted gene following Bal31 treatment. This was not possible with the inserted *dnaB* gene fragment utilized in the construction of the *dnaB* overproducer pPS359. In this case the lack of an appropriate restriction endonuclease site downstream of the gene meant that in the final product plasmid there remained a considerable length of DNA whose sequence had not previously been determined. To elucidate the entire nucleotide sequence and size of pPS359 it was necessary to sequence this remaining DNA.

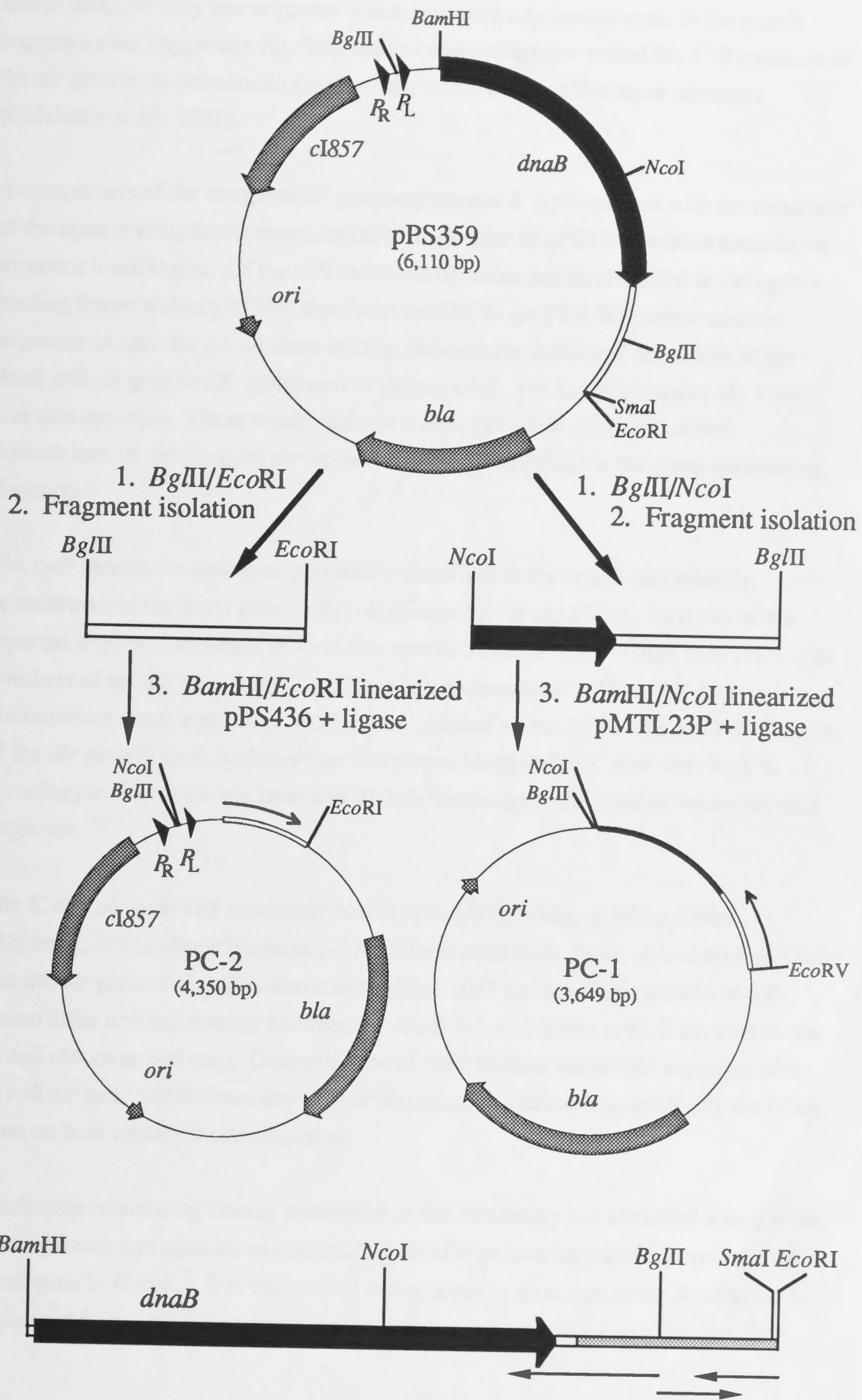
Much of this DNA was sequenced in one direction by directly sequencing alkali-denatured pPS359 plasmid DNA, using appropriate primers (Figure 2.1).

Determination of the remaining sequence between this and the predicted position of the translation termination codon of the *dnaB* gene required digesting the plasmid with a restriction endonuclease that would cleave the unsequenced region of DNA to provide templates for further sequencing reactions. The two *Bgl*II fragments isolated from pPS359 (Figure 4.14) were used to obtain the complete sequence of one strand in this region of the plasmid, and led to identification of the 3'-OH terminus of the *dnaB* gene and the site of fusion with the parent vector downstream of the *dnaB* gene. The sequence is given in Appendix A.

Analysis of the newly sequenced region of DNA in pPS359 revealed the 5' end of a potential open reading frame 55 bp downstream of the translation termination (tandem ochre) codons of the *dnaB* gene. The open reading frame commences with an ATG translation initiation codon and is not terminated within the chromosomal DNA fragment in plasmid pPS359. Using the Genbank genetic sequence data bank, a search was instigated for homologies with the first 50 ntds of this open reading frame (a sequence determined by this laboratory and by Nakayama *et al.*, 1984b). This

Figure 4.14

Sequencing of DNA downstream of *dnaB* in pPS359. Over 200-ntds of DNA was sequenced directly from the plasmid using primer 10 (Figure 2.1). The remaining DNA was sequenced in either direction from a *Bgl*III site downstream of the *dnaB* gene following the construction of two plasmids (PC-1 and PC-2) containing the relevant DNA fragments. Arrows indicate the direction of sequencing and the filled box in PC-1 represents the remaining 3' end of *dnaB* downstream of the *Nco*I site of the gene. The *Bgl*III site used in screening of plasmids for anticipated size and for fragment inserts is centred 278-bp upstream of the *Bam*HI site in pPS436 (Figure 4.9). The direction and extent of sequencing from these reactions is demonstrated with the *Bam*HI/*Eco*RI fragment containing the *dnaB* gene and that DNA previously unsequenced (shaded box).



search revealed only one sequence which was perfectly homologous to the search sequence used (Appendix A). The sequence identified comprised the 5'-P terminus of the *alr* gene from *Salmonella typhimurium* which encodes D-alanine racemase (Galakatos *et al.*, 1986).

A comparison of the complete *alr* gene sequence of *S. typhimurium* with the remainder of the open reading frame downstream of *dnaB* gene in pPS359 revealed remarkable sequence homologies. Of the 630 ntds (or 210 amino acids) identified in the open reading frame within pPS359, there was an 82.7 % (or 92.4 %) conservation of sequence (Appendix A). A short overlap between the published sequences of the *dnaB* and *alr* genes of *S. typhimurium* (Wong *et al.*, 1988; Galakatos *et al.*, 1986) was also apparent. These results suggest that *E. coli alr* is also immediately downstream of *dnaB* on the chromosome, and is transcribed in the same (clockwise) direction.

The *tyrB* gene is the next gene previously identified in the region immediately downstream of the *dnaB* gene in *E. coli* (Kuramitsu *et al.*, 1985). Analysis of the reported sequence upstream of *tyrB* also revealed extensive homology with the 3'-OH terminus of the *alr* gene from *S. typhimurium* (Appendix A). With this sequence and the sequence determined in pPS359, homologies of all but 337 of the 1077 nucleotides of the *alr* gene from *S. typhimurium* have been identified in *E. coli* with 81.8 % homology at the nucleotide level and 92.8 % homology at the level of the amino acid sequence.

The *E. coli alr* gene had previously been mapped imprecisely at ~93 minutes (Wijsman, 1972), about 1 minute (~470 kb) removed from *dnaB*. It is therefore likely that the *alr* gene of *E. coli* is also comprised of 1077 bp, encodes a protein of 359 amino acids and lies directly between the *dnaB* and *tyrB* genes at 91.9 minutes on the *E. coli* chromosomal map. Determination of the complete nucleotide sequence of the *E. coli alr* gene and surrounding genetic elements by dideoxy sequencing of the DNA gene on both strands is contemplated.

Nucleotide sequencing already performed in this laboratory has identified a long open reading frame that encodes an unidentified 38-kDa protein immediately upstream of the *dnaB* gene in *E. coli*.¹ It is transcribed in the opposite direction to the *dnaB* gene, is expressed from a plasmid containing the gene, and is homologous to the gene that

1. P.E. Lilley and N.E. Dixon, unpublished.

occurs in the same position in *S. typhimurium* (Wong *et al.*, 1988). This suggests that the *dnaB* gene, at least, has its own promoter(s) for transcription. There are a number of candidate sequences that resemble promoter sites for RNA polymerase (Appendix A). The only extensive dyad symmetry that may be associated with transcription termination was found downstream of the *alr* gene in both *E. coli* and *S. typhimurium* sequences (Appendix A). In both cases the GC-rich palindrome was followed by a stretch of consecutive AT base pairs, which is characteristic of a *rho*-independent transcription terminator (Rosenberg and Court, 1979). Considering that the intergenic region between *alr* and *dnaB* in both *E. coli* and *S. typhimurium* sequences is relatively AT rich, it is likely that the *dnaB* and *alr* genes of *E. coli* and *S. typhimurium* are cotranscribed as an operon. Further work is necessary to test this hypothesis.

DISCUSSION

Overproduction of dnaB, dnaC and dnaT proteins

Like primase (Chapter 3), the high-level expression of the *dnaB*, *dnaC* and *dnaT* genes was achieved by replacing the sequences responsible for the normal control of their expression. In each case, the natural promoter region and/or ribosome-binding site was removed by treatment of an appropriate DNA fragment with the double-stranded exonuclease Bal31. Alternate strategies were then employed to: (i) provide a new ribosome-binding site perfectly complementary to the 3'-OH terminus of *E. coli* 16-S ribosomal RNA, variably spaced upstream of the ATG translation initiation codon; and (ii) to place transcription of the gene under control of tandem strong bacteriophage λ promoters P_R and P_L in vector derivatives of pCE30 (Elvin *et al.*, 1990).

The main objective in engineering plasmids to direct the overexpression of the *dnaB*, *dnaC* and *dnaT* proteins was the recovery of these proteins in quantities sufficient for chemical and structural studies. To this end, the most crucial step in a large-scale purification of these proteins would be the recovery of soluble and active protein in lysates, since this would ultimately influence the total quantity of the protein recovered. However, unlike primase, the proteins overproduced in these strains were almost completely insoluble following all lysis procedures that were investigated.

It has been well documented that in *E. coli*, systems designed to express high levels of proteins from cloned genes often produce protein which accumulates intracellularly in an insoluble form (Kane and Hartley, 1988). The protein in these proteinaceous aggregates is usually denatured and purification of biologically active, soluble protein from such aggregates requires solubilization of the protein and its refolding to its native and active form (Marston and Hartley, 1990). Due to the insolubility of the starting material, such a strategy leading to recovery of the overproduced protein has the advantage of providing a useful means of increasing product purity by immediate purification from the bulk of the soluble proteins. Unfortunately, refolding the polypeptide chain *in vitro* does not guarantee the recovery of full biological activity. Such a loss may occur on account of (i) a loss of conformational authenticity in the refolded protein, (ii) omission of pertinent co-factors normally incorporated into the apo-enzyme, (iii) modifications resulting from the solubilization and renaturation conditions, or (iv) loss of post-translational modifications.

For these reasons, solubilizing and refolding the proteins from the insoluble lysate fraction was not considered an appropriate method for the purification of the dnaB, dnaC and dnaT proteins. The alternative was to utilize other techniques to direct overexpression of these proteins in an environment conducive to folding the nascent linear polypeptide into its native and biologically-active conformation *in vivo*. Two options were investigated utilizing the dnaB overexpression vector pPS359. Although overproduction of the protein at a lowered temperature by chemical induction had little or no affect on solubility, the co-expression of dnaB with dnaC protein in the same cell led to solubilization of both proteins. This suggests that a main factor influencing the folding of these proteins into their native conformation was not one of temperature dependence but concentration, where the presence of one protein is required in more or less stoichiometric amounts for the folding and solubility of the other.

To be biologically active, all proteins must adopt specific folded three-dimensional structures, and the folding of many proteins *in vitro* can be described as a two step process (Creighton, 1990). Conversion of the unfolded protein (U) to an intermediate (I) involves the formation of secondary structural elements. In the second (rate-determining) step, the organization of these elements into specific tertiary structures associated with the native state (N) occurs.



Solvent has greater accessibility to non-polar groups in the I state(s) than in the N state. Consequently, the I state is characteristically much less stable than the N state and is susceptible to aggregation. The efficiency of protein folding is therefore dependent on the kinetic partitioning of the I state(s) into the N state or the aggregated state, I_{agg} .



The formation of inclusion bodies in cells overexpressing recombinant proteins is a manifestation of partitioning to the aggregated state.

The genetic information for the protein specifies directly only the primary structure, the linear sequence of the amino acids in the polypeptide backbone. The model for protein folding presented here is dependent on the assumption that an unfolded linear chain of amino acids is able to interact with itself and with its physiological environment to assume a conformation of lower free energy, the principle of self assembly (Anfinsen, 1973). However, many polypeptides assemble only in the presence of additional proteins (Ellis and Hemmingsen, 1989; Rothman, 1989). These polypeptide chain binding proteins, 'molecular chaperones' or 'chaperonins', mediate the folding of certain other polypeptides, and, in some instances, their assembly into oligomeric structures. They may do this primarily by preventing the formation of certain intermediate structures that might otherwise occur and lead to a non-functional conformation.

Although in cases where folding is not mediated by chaperones, the stability and solubility of many overproduced *E. coli* multi-subunit proteins may still be directly attributed to similar protein-protein interactions. An excellent example of this phenomenon was observed with *E. coli* integration host factor (IHF), a protein composed of two non-identical polypeptides IHF- α and IHF- β (Nash *et al.*, 1987). Separate expression of these proteins under the control of the λP_L promoter led to the production of unstable and insoluble peptides. In contrast, the overexpression of both genes conjointly led to the accumulation of large amounts of active IHF. Similarly, simultaneous overproduction of subunits of other multi-subunit proteins has led to the expression of stable and soluble proteins (Kane and Hartley, 1988). In each of these cases it is possible that the close association of the subunits of these complexes directly contributes to the partitioning of the protein into the native and biologically active conformation (N). Thus each subunit effectively acts as a 'molecular chaperone' for the other.

Such a partitioning was observed for the overproduction of the *dnaB* and *dnaC* proteins. In this case, the separate overexpression of each gene was not a usable method for overproducing soluble protein products. Neither the normal association of *dnaB* protomers nor overproduction at a lowered temperature was able to confer stability to the hexameric *dnaB* protein. Interestingly, however, the presence of approximately stoichiometric amounts of *dnaC* appeared to solubilize intracellular *dnaB* protein and the presence of stoichiometric amounts of *dnaB* appears to stabilize intracellular *dnaC* protein. The *dnaB* and *dnaC* proteins have been shown to form a complex when mixed *in vitro* in the presence of ATP (Kobori and Kornberg, 1982c). The dissociation constant (K_d) for this complex suggests that the two proteins, at normal concentration *in vivo*, probably occur in the *dnaB*-*dnaC* complex state (Wahle *et al.*, 1989a). It is quite likely, therefore, that each protein acts as a 'chaperonin' to direct the kinetic partitioning of the other protein into the native state (N) rather than the aggregated state (I_{agg}).

Transposing this model to the intracellular environment of wild-type *E. coli* cells, the conformational authenticity and biological activity of both *dnaB* and *dnaC* proteins may also be dependent on the presence of the other protein in approximately equimolar amounts. This requirement may therefore provide a final regulatory mechanism directing the total quantity of each species normally present *in vivo*. A failure to regulate low-level production of one protein would be balanced by the scarcity of the other (and therefore not affect the viability of the cell).

Whatever the intricacies involved in *dnaB* and *dnaC* protein association and regulation in wild-type *E. coli*, this finding provides further evidence that the two proteins interact directly *in vivo*. It also offers a cautionary note to workers attempting to overproduce other proteins that are either composed of more than one subunit or proteins intimately associated under normal conditions *in vivo*. The use of the dual promoter vector pPS436 may, in these cases, prove an invaluable tool in the overproduction of both subunits of heterodimeric proteins. An immediate target for this overproduction strategy would be the *dnaT* gene. Separate expression of this gene under the λ promoters P_R and P_L , like the *dnaB* and *dnaC* genes, led to the production of an insoluble product. The *dnaT* protein, like *dnaB*, naturally occurs as a multi-subunit complex of identical subunits, in this case a trimer (Arai *et al.*, 1981g). However, the association of these subunits obviously has little effect on influencing the production of a stable *dnaT* conformation. It is possible that concomitant expression of the *dnaT* gene with another protein, possibly one of primosomal origin such as *dnaB* or a bacterial chaperonin such as *groEL* or *dnaK* protein, may direct the partitioning of the nascent *dnaT* polypeptide into its native and soluble conformation.

These and similar experiments directed at engineering high-level expression of the *dnaB* protein with the *E. coli* chaperone *groEL* and the P protein of phage λ (also insoluble when overproduced; Tsurimoto *et al.*, 1982) are being contemplated.

Separate purification of the *dnaB* and *dnaC* proteins and the *dnaB*-*dnaC* complex, in quantities adequate for structural studies, is described in the following chapter (Chapter 5).

The dnaB-alr operon of E. coli

Sequence determination of DNA downstream of the *dnaB* gene, to elucidate the entire nucleotide sequence and size of pPS359, revealed the partial sequence of another open reading frame (*alr*) encoding the protein alanine racemase. The absence of likely promoter sequences immediately preceding *alr* and possible *rho*-independent termination sequence downstream of *alr*, suggest that the *dnaB* and *alr* genes form an operon. The alanine racemase protein of *Salmonella typhimurium* catalyses the interconversion of L- and D-alanine, providing the latter enantiomer for the construction of the peptidoglycan layer of the bacterial cell wall (Esaki and Walsh, 1986). Thus the *dnaB* and *alr* genes may form a macromolecular synthesis operon coupling expression of genes required for DNA replication and cell wall synthesis (Figure 4.15; Appendix A). The complete nucleotide sequence of the *E. coli alr* gene and surrounding genetic elements will be determined in future work.

Several other DNA replication genes have also been found associated at the level of gene regulation with other macromolecular processes in *E. coli*. These include the *dnaG* gene encoding primase, the *dnaE* gene encoding the α subunit of DNA polymerase III holoenzyme and the *priB* gene encoding primosomal protein PriB. The *dnaG* gene forms an operon with two genes, *rpsU* and *rpoD*, involved in translation and transcription respectively (Lupski and Godson, 1984; see Chapter 3, Figure 3.1). The *priB* gene is in an operon containing genes encoding ribosomal proteins S6 (*rpsF*), S18 (*rpsR*) and L9 (*rplI*) (Schnier *et al.*, 1986; Allen and Kornberg, 1991; Zavitz *et al.*, 1991). The *dnaE* gene, moreover, lies in a complex operon that contains at least seven genes (Figure 4.16), including two genes required for lipid A biosynthesis and another gene encoding RNaseHII (Coleman and Raetz, 1988; Itaya, 1990). Lipid A provides the hydrophobic anchor for lipopolysaccharide, the major component of the outer monolayer of the cell's outer membrane, while RNaseHII specifically degrades the ribonucleotide moiety in RNA-DNA hybrid molecules and

Figure 4.15

The predicted arrangement of the genes and regulatory units of the *dnaB-alr* operon at 92 minutes on the *E. coli* chromosome. Arrows indicate the direction of the transcription and the numbers below the line represent in base pairs the sizes of the genes and regions between the genes and the predicted regulatory units. The complete length of the *alr* gene in *E. coli* was deduced from the published sequence of *E. coli tyrB*, the homologous *S. typhimurium alr* gene, and the sequence of pPS359 (Appendix A). An open reading frame upstream of the *dnaB* gene in *E. coli* encoding a 38-kDa protein has been identified (Dr N.E. Dixon, personal communication) and its direction of transcription suggests that the *dnaB* protein is transcribed from its own promoter sequences. A GC-rich palindrome followed by a long stretch of consecutive AT base pairs downstream of the *alr* gene is characteristic of *rho*-independent transcription termination. This potential transcription termination site may provide the terminator for transcripts initiated by the *dnaB* promoter sequences.

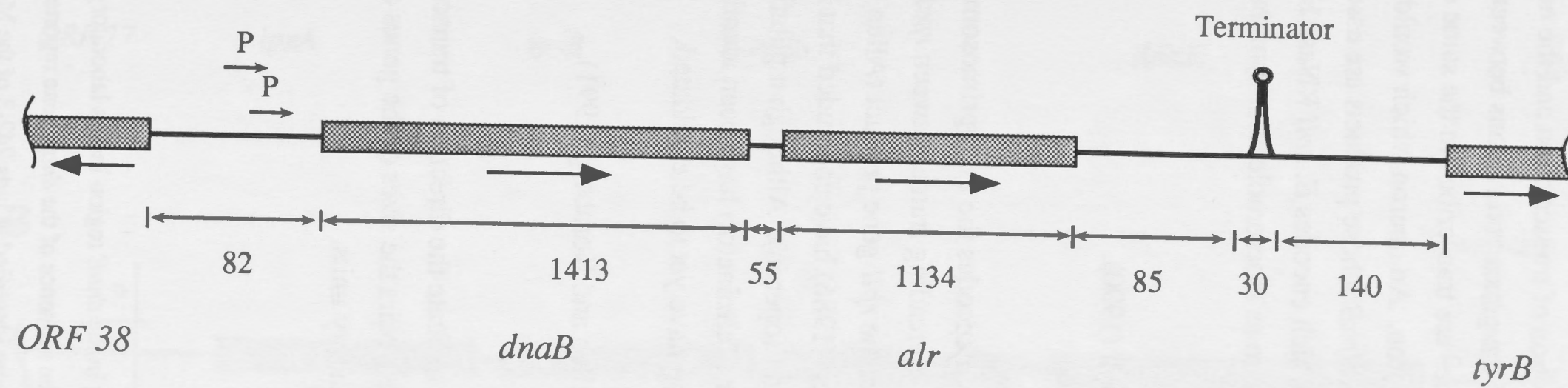


Figure 4.16

A. Gene location and orientation of the 4.5 minute region of the *E. coli* chromosome. Arrows indicate the direction of transcription and the numbers below the line represent in base pairs the sizes of the genes and regions between the genes. Since the genes adjacent to the *dnaE* gene² are transcribed in the same direction, it is possible they reside on a common operon. An operon which would consist of at least six other genes including *lpxA* and *lpxB* whose products are essential for lipid A disaccharide biosynthesis and *rnhB* which encodes *E. coli* RNase HII (Itaya, 1990). This putative operon has been referred to as macromolecular synthesis II (Tomasiewicz and McHenry, 1987).

Figure adapted from Itaya (1990).

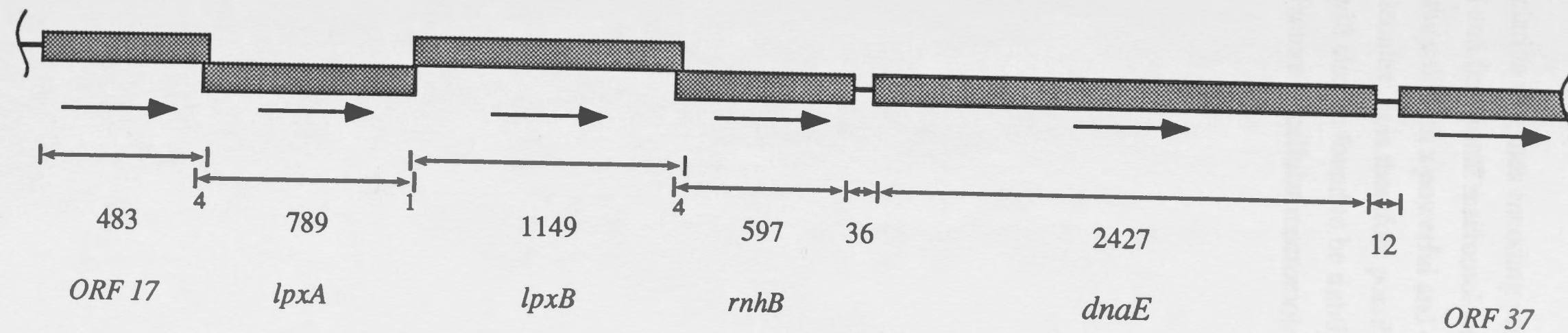
B. The *priB* gene, which encodes the PriB primosomal protein, corresponds to a previously identified open reading frame between *rpsF* and *rpsR* within a ribosomal operon that also encodes the *rplI* gene product (Allen and Kornberg, 1991; Zavitz *et al.*, 1991). Schnier *et al.* (1986) have concluded that *rpsF-priB-rpsR-rplI* form a single transcriptional unit (operon). Although a putative promoter (P) and transcription terminator (Terminator) have been identified, details of the transcriptional regulation of this operon have yet to be elucidated.

Figure adapted from Allen and Kornberg (1991).

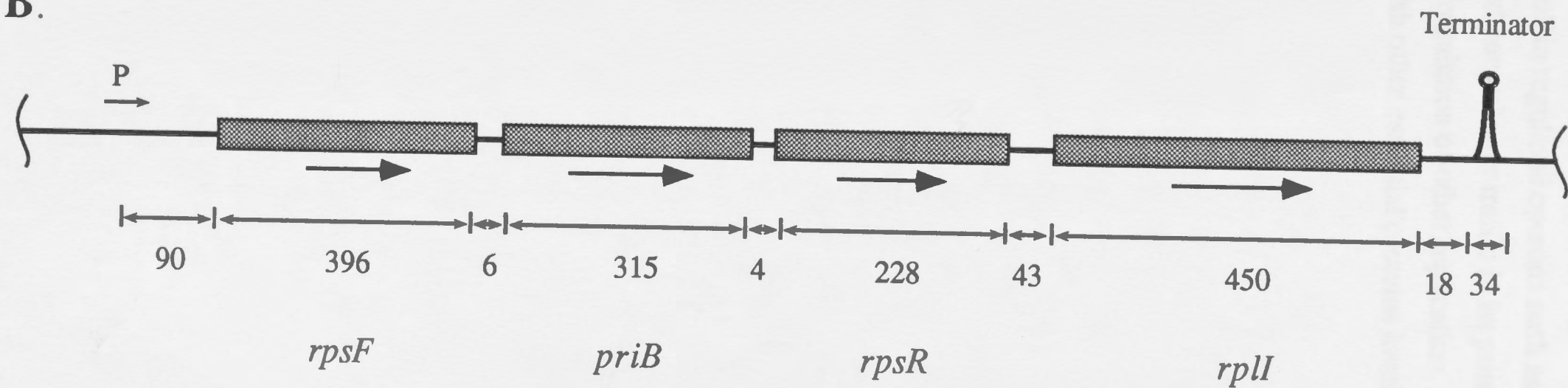
In both figures arrows indicate the direction of transcription and the numbers below the line represent in base pairs the sizes of the genes and regions between the genes and the predicted regulatory units.

2. Nucleotide sequencing in the *dnaE* region in this laboratory (N.E. Dixon, unpublished) has identified a conflict with the sequence of the *dnaE* gene reported by Tomasiewicz and McHenry (1987). Only one T and not two were identified at ntids 742-3 of the McHenry sequence. This changes the open reading frame of *rnhB* so that it terminates at a TGA 36 ntids upstream of the *dnaE* ATG translation initiation codon and produces a protein 14 amino acids shorter than previously reported.

A.



B.



may be implicated in several processes in DNA replication, in DNA repair, and in SOS-induced mutagenesis.

The association of genes encoding proteins essential for the biosynthesis of the major structural and functional macromolecules in *E. coli* in regulated operons such as these provides the cell with a powerful and co-ordinated control over traffic in its primary macromolecules. It is therefore possible that the regulation of other replication proteins will also be found to be tightly linked with other essential proteins involved in the manufacture of cellular macromolecules.

CHAPTER 5

Purification and Properties of the *dnaB* and *dnaC* Proteins, and a *dnaB*-*dnaC* Complex

The manipulation of several well defined *in vitro* DNA replication systems has demonstrated that dnaB protein is a crucial replication enzyme, essential in the initiation and elongation stages of replication of the *E. coli* chromosome. The dnaB protein has two known functions in DNA replication: (i) it enables *E. coli* primase to repeatedly generate short RNA transcripts needed to prime synthesis of nascent fragments of a lagging strand at the replication fork (Arai and Kornberg, 1981e; Arai *et al.*, 1981f; LeBowitz *et al.*, 1985); and (ii) as a helicase it separates the DNA strands in front of DNA polymerase (LeBowitz and McMacken, 1986; Baker *et al.*, 1986).

E. coli DNA replication mutants exhibit two diverse phenotypes, those that are thermosensitive for fork movement (fast stop) and those that block replication initiation (slow stop). With the dnaC protein, another key replication enzyme, both slow and fast stop *dnaC^{ts}* mutants have been isolated (Wechsler and Gross, 1971; Wechsler, 1975). These indicate that dnaC protein may function in the initiation of DNA replication at the chromosomal origin as well as in elongation, although no function of dnaC that does not involve interaction with dnaB protein has been identified. Genetic evidence (Section 4.3) implicates an interaction between these two proteins *in vivo* and association of the purified proteins in the presence of ATP or dATP directs the formation of a tight, isolable complex in which the proteins occur in 1:1 molar ratio (Wickner and Hurwitz, 1975; Kobori and Kornberg, 1982c; Lanka and Schuster, 1983; Wahle *et al.*, 1989a).

The successful concurrent overproduction of soluble dnaB and dnaC proteins in *E. coli* (Chapter 4) enables the purification of sufficient quantities of these proteins for close examination of their physiochemical properties, mechanisms of action and structures. Of particular interest was the isolation of an intact dnaB-dnaC complex. Kobori and Kornberg (1982c) reported that following mixing of purified dnaB and dnaC protein in the presence of ATP, the free dnaC and dnaB proteins and a dnaB-dnaC complex could be isolated separately by anion-exchange chromatography on a DEAE-cellulose column. Separation was also observed by Arai *et al.* (1981b) and Lanka and Schuster (1983) when purifying the dnaB protein and dnaC protein, respectively. In each case some dnaB protein activity coeluted from a DEAE-cellulose column at a low salt concentration and coincided with a peak of dnaC activity, suggesting an interaction between the dnaB and dnaC proteins.

During purification, dnaC protein normally flows through a DEAE-cellulose column (Kobori and Kornberg, 1982b; Lanka and Schuster, 1983); dnaB protein binds and elutes following the application of a salt gradient (Reha-Krantz and Hurwitz, 1978;

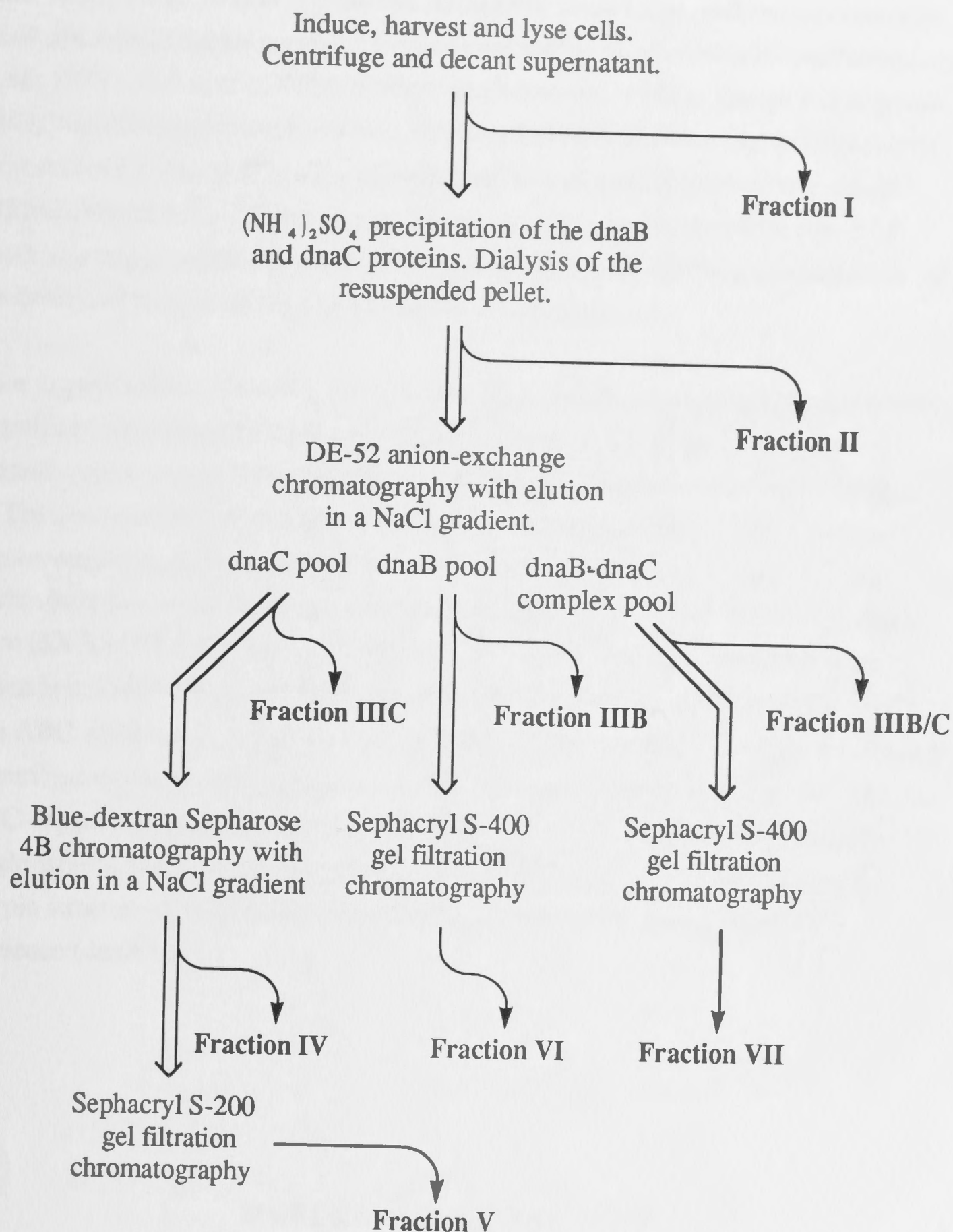


Figure 5.1

Strategy for the separation and isolation of the dnaB and dnaC proteins and the dnaB-dnaC complex from cells harvested from a thermoinduced culture of RSC680 (pPS562; Figure 4.11) or RSC661 (pPS554; Figure 4.10). The three species in a cell-free extract are separated by use of DEAE anion-exchange chromatography.

Ueda *et al.*, 1978; Arai *et al.*, 1981b). Using this knowledge, and the fact that both dnaB and dnaC proteins precipitate at low concentrations of ammonium sulfate (Ueda *et al.*, 1978; Arai *et al.*, 1981b; Kobori and Kornberg, 1982b; Lanka and Schuster, 1983), the strategy reported here was designed for the separation and purification of proteins overproduced following thermal induction of cells containing the plasmid pPS562 (Figure 5.1). This protocol allowed not only the removal of the bulk of contaminating proteins, but also enabled the simultaneous and complete separation of the dnaB and dnaC proteins and a dnaB-dnaC complex.

Prior to purification, however, it was necessary to develop an appropriate method for quantification of the enzymatic activity of dnaB and dnaC. Three appropriate reconstituted assay systems dependant on both dnaB and dnaC have been described: (i) The *dnaA*-dependant replication of double-stranded templates containing the chromosomal origin of replication, *oriC* (Kaguni and Kornberg, 1984); (ii) The primosome-dependent conversion of single stranded ϕ X174 DNA to the replicative form (ϕ X SSARF; Shlomai *et al.*, 1981); and (iii) The recently described dnaA dependent SSARF reaction with single-stranded DNA containing the R6K γ 2 origin (the ABC primosome assay; Masai *et al.*, 1990b). In consideration of the availability of purified proteins, the last assay was most accessible for use in this work. The ABC-dependent priming mechanism operates on a single-stranded DNA coated with single-stranded binding protein (Masai *et al.*, 1990b) and is initiated at a specific hairpin structure in the template whose stem carries a dnaA protein recognition sequence (dnaA box).

MATERIALS AND METHODS

5.2.1 dnaB - dnaC Dependent ABC Assay

Sources of *E. coli* replication proteins were: Highly purified single-stranded DNA binding protein (8.8×10^4 units/mg)¹ and the β subunit of polymerase III holoenzyme (3×10^6 units/mg)² were gifts from Dr P. Hendry and Dr J.L. Beck (Research School of Chemistry, Australian National University). Partially-purified DNA polymerase III* (5.5×10^5 units/mg), purified through Fraction IV (Maki *et al.*, 1988), and dnaA protein (3.1×10^5 units/mg)³ were gifts from Dr N.E. Dixon (Research School of

Chemistry, Australian National University). The *dnaG* primase (1.36×10^6 units/mg) was as described (Section 3.2.3). The *dnaC* (2.1×10^5 units/mg) and *dnaB* (1.2×10^5 units/mg) proteins were prepared as described below (Section 5.2.2). The *ssiA*(R6K γ 2) template (Masai *et al.*, 1990b) was a gift from Dr H. Masai (DNAX Research Institute of Molecular and Cellular Biology).

Reaction mixtures (25 μ l) for the ABC-priming assay SS-DNA replication contained: 20 mM Tris.HCl pH 8.0, glycerol (4 % v/v), dithiothreitol (8 mM), bovine serum albumin (80 μ g/ml), magnesium acetate (8 mM), ATP (2 mM), CTP, GTP and UTP (100 μ M each), dATP, dCTP and dGTP (50 μ M each), [3 H]-TTP (50 μ M, 3.3 Bq/pmol), *ssiA*(R6K γ 2) single-stranded DNA template (230 pmol, as nucleotide), single-stranded DNA binding protein (600 ng), DNA polymerase III* (100 units) and the β subunit (16 ng), *dnaA* protein (62 ng), *dnaG* primase (87.5 ng), and when quantifying *dnaB* or *dnaC* activity separately, either *dnaB* protein (130 ng) or *dnaC* protein (84 ng). The mixtures were assembled at 0 °C then, after addition of *dnaB* protein, *dnaC* protein or the *dnaB*-*dnaC* complex, were treated for 10 min at 30 °C. DNA synthesis was terminated by cooling at 0 °C and addition of trichloroacetic acid (10 % w/v containing 0.1 M NaPP_i). Acid precipitates were collected on Whatman GF/C filter disks which were washed with 1 N HCl, 0.1 M NaPP_i and ethanol, dried and counted in a liquid scintillation counter. One unit of enzymatic activity is the amount that promotes incorporation of one pmol of nucleotide into product in one minute.

5.2.2 Purification of the *dnaB* and *dnaC* Proteins and the *dnaB*-*dnaC* Complex from RSC680

The final composition of the lysis buffer was: 50 mM Tris.HCl pH 7.6, 10 % (w/v) sucrose, 0.1 M NaCl, 2 mM dithiothreitol, 10 mM spermidine.HCl. Buffers used for chromatography were 50 mM Tris.HCl pH 7.6, 20 % (v/v) glycerol, 5 mM MgCl₂, 2 mM dithiothreitol and 100 μ M ATP containing either 25 mM NaCl (Buffer A) or 200 mM NaCl (Buffer B); 50 mM Tris HCl, pH 8.2, 20 % (v/v) glycerol, 4 mM dithiothreitol and 1 mM EDTA containing either 25 mM NaCl (Buffer C) or 200 mM NaCl (Buffer D).

-
1. P. Hendry and N.E. Dixon, unpublished.
 2. J.L. Beck and N.E. Dixon, unpublished.
 3. N.E. Dixon, unpublished.

Strain RSC680 (containing plasmid pPS562) was grown at 30 °C in LB broth containing 50 mg/l ampicillin in 5 x 1-l baffled flasks to $A_{595} = 0.5$. After rapid temperature shift to 42 °C, growth was continued at 42 °C for a further period of four hours (final $A_{595} = 1.56$). The cultures were chilled; cells (7.17 g wet weight) were harvested (8,000 x g, 15 min), resuspended in 29.3 ml of 50 mM Tris.HCl pH 7.6, 10 % (w/v) sucrose, poured into liquid N₂ and stored at -70 °C.

The RSC680 cell suspension was thawed and diluted to 313 ml to the final composition of lysis buffer. Lysozyme (0.2 mg/ml) was added and the mixture was stirred for 1 h at 0 °C and treated at 37 °C for 4 min, inverting the mixture every 1 min to lyse cells before returning to ice. To the supernatant (Fraction I, 305 ml) obtained after centrifugation (12,000 x g, 20 min) was added solid ammonium sulfate (0.2 g/ml). After being stirred for 1 h, the suspension was centrifuged (40,000 x g, 45 min). The pellet was dissolved in 30 ml of Buffer A and dialysed for 2 h vs two changes (1-l each) of Buffer A and diluted to 100 mls with this buffer (Fraction II). Fraction II was applied to a column (16.3 x 2.5 cm) of DEAE-cellulose (DE-52) anion-exchange resin pre-equilibrated in Buffer A. The column was washed with Buffer A (105 ml) and bound proteins were eluted with a linear gradient (800 ml) of 0.025 to 0.6 M NaCl in Buffer A at a flow rate of 45 ml/h.

Fractions containing uncomplexed dnaB protein (Fraction IIIB, 102 ml) and the dnaB-dnaC complex (Fraction IIIB/C, 82.5 ml) were pooled and stored at -70 °C (see Figure 5.4). Uncomplexed dnaC protein not bound to the column was pooled and dialysed for 2 h vs two changes (1-l each) of Buffer C (Fraction IIIC, 155 ml). Fraction IIIC was applied to a column (10.2 x 2.5 cm) of blue dextran-Sepharose resin pre-equilibrated in Buffer C. The column was washed with Buffer C (50 ml); bound dnaC protein was eluted with a linear gradient (400 mls) of 0.025 to 2 M NaCl in Buffer C at a flow rate of 48 ml/h. The bound dnaC protein eluted in a single peak coincident with the single peak of protein, and was pooled (Fraction IV, 53.5 ml).

Solid ammonium sulfate (0.35 g/ml) was added to Fraction IV and the suspension was stirred for 1 h. The pellet obtained after centrifugation (40,000 x g, 45 min) was dissolved in 3 ml of Buffer D and dialysed for 2 h vs 1 l of this buffer. The concentrated Fraction IV was diluted to 4 mls in Buffer D and applied to a column (2.5 x 44 cm) of Sephacryl S-200 gel filtration resin pre-equilibrated with this Buffer. The protein was eluted at a flow rate of 45 ml/h and the dnaC-active fractions were pooled (Fraction V, 62 ml).

To Fractions IIIB and IIIB/C solid ammonium sulfate (0.25 and 0.3 g/ml respectively)

was added. After being stirred for 1 h, the suspensions were centrifuged ($40,000 \times g$, 45 min), and the pellets dissolved in 3 ml of Buffer B and dialysed for 2 h vs 1 l of this buffer. At this point protein from Fractions IIIB and IIIC were treated in identical fashion. Each was diluted to 4 ml with Buffer B and loaded onto a column (2.5 x 44 cm) of Sephacryl S-400 gel filtration resin pre-equilibrated with this buffer. The proteins were eluted at a flow rate of 45 ml/h and the fraction containing dnaB (Fraction VI, 61 ml) or dnaB-dnaC (Fraction VII, 38 ml) were pooled. Fractions V (dnaC), VI (dnaB) and VII (dnaB-dnaC) were stored at -70°C until required.

RESULTS

ABC-Priming Assay

During purification the activity of the dnaB and dnaC proteins was assessed using the dnaA-dependent SS \rightarrow RF conversion of *ssiA*(R6K γ 2) single-stranded DNA template (Masai *et al.*, 1990b). In developing use of this novel mode of priming for replication assays, efficiency of replication (relative to Masai *et al.*, 1990b) was increased by halving the reaction volumes, substituting the buffer system used to that of the G4 SS \rightarrow RF assay (Section 3.2.1), and using preparations of highly-purified proteins. The result was a perfectly reproducible assay with an increase in efficiency ~ 4 -fold over that previously reported (Table 5.1). Addition of rifampicin to suppress RNA polymerase-dependent priming was not necessary as evidenced by the low level of DNA synthesis associated with the primase-free reaction.

The specific activity of both dnaA protein and the dnaG primase in promoting DNA replication in this assay were comparable to the activities of these proteins assessed by other replication assays. The specific activity of highly-purified primase (1.36×10^6 units/mg; Figure 5.2A) was identical to that measured in the G4 SS \rightarrow RF assay (Chapter 3) and the specific activity of the partially-purified dnaA protein (3.51×10^5 units/mg; Figure 5.2D) was close to that measured in a dnaA^{ts} mutant complementation assay (3.07×10^5 units/mg; Dr N.E. Dixon, unpublished results) with M13oriC26RF DNA as template (Fuller and Kornberg, 1983). This makes ABC-priming on this template an excellent assay for either protein. The specific activities of the highly-purified dnaB and dnaC proteins were calculated as 1.2×10^5

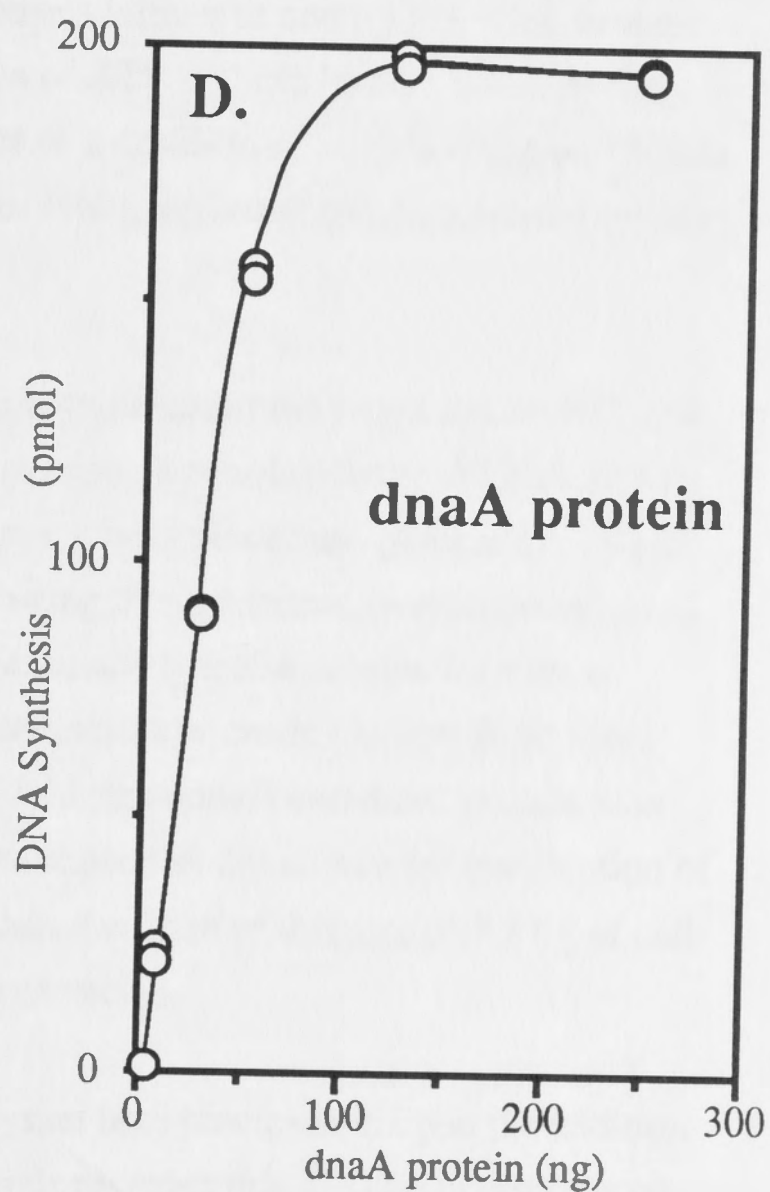
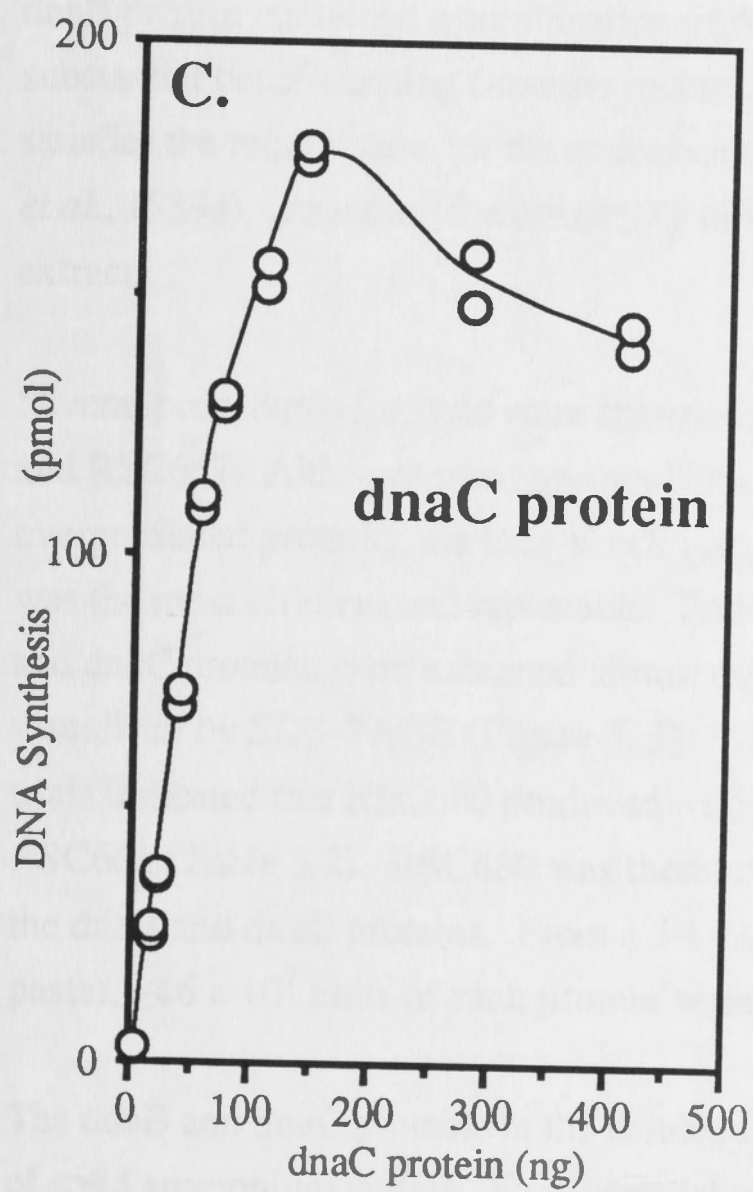
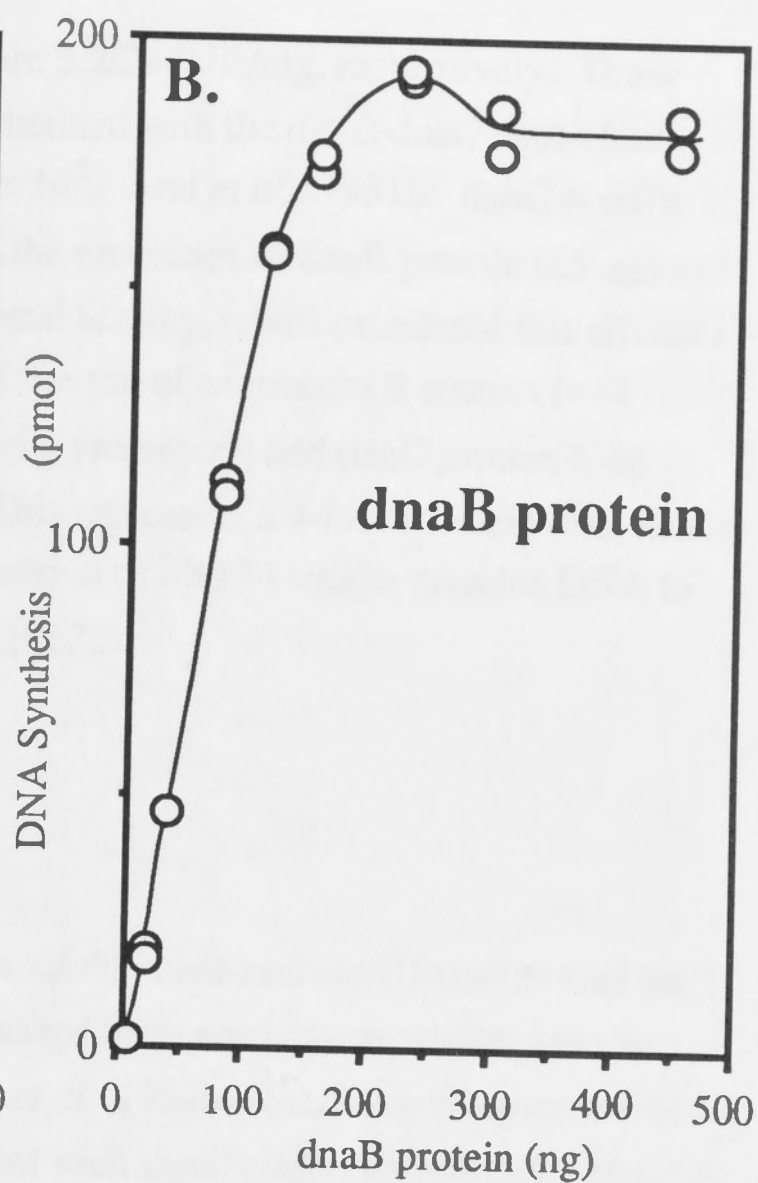
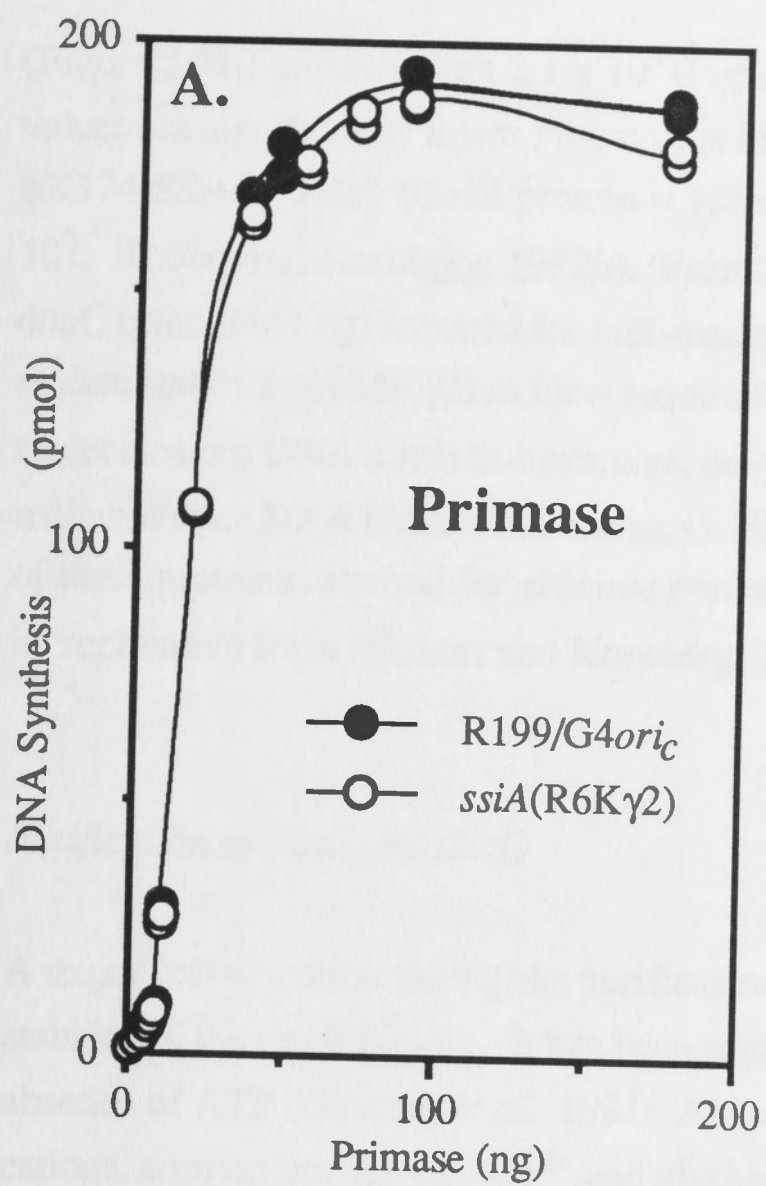
Table 5.1

The ABC Primosome Assay. Protein requirements for reconstituted SS→RF replication of *ssiA*(R6Kγ2) DNA. The standard SS→RF DNA replication reaction (10 min at 30 °C) was performed as described in Section 5.2.1 with 230 pmol (as deoxyribonucleotide) of *ssiA*(R6Kγ2) single-stranded DNA template. Values are compared to those obtained by Masai *et al* (1990b), who used 300 pmol of template at 30 °C over 10 min.

Components Omitted	DNA Synthesis	
	(Masai <i>et al.</i> , 1990b)	(Section 5.2.1)
	<i>pmol</i>	<i>pmol</i>
None	72	199
dnaA protein	2	2
dnaB protein	<1	1
dnaC protein	2	2
Primase	8	3
DNA polymerase III holoenzyme	<1	<1
SSB and dnaA protein	98	166
SSB, dnaA protein and dnaC protein	94	127

Figure 5.2

Requirement for the primosomal proteins dnaB, dnaC and primase, and the dnaA protein in the *ssiA*(R6K γ 2) SS \rightarrow RF conversion. A, requirement of primase (Fraction IV; Section 3.2.3) in total reconstitution reactions with *ssiA*(R6K γ 2) (Section 5.2.1) and R199/*G4oriC* (Section 3.2.2) single-stranded DNA templates (230 pmol). Assays were at 30 °C for 10 min. Specific activities for primase in the *ssiA*(R6K γ 2) and R199/*G4oriC* SS \rightarrow RF reactions were both calculated as 1.36×10^6 units/mg. B, requirement of dnaB protein (Fraction VI; Section 5.2.2), C, requirement of dnaC protein (Fraction V; Section 5.2.2), and D, requirement of dnaA protein, in total reconstitution reactions as described (Section 5.2.1). The specific activity of partially-purified dnaA protein (3.51×10^5 units/mg) is comparable to the specific activity of this protein determined using the *oriC* dnaA^{ts} mutant complementation assay (3.07×10^5 units/mg; N.E. Dixon, unpublished results).



(Figure 5.2B) units/mg and 2.1×10^5 (Figure 5.2C) units/mg, respectively. These values are significantly lower than results obtained with the dnaB-dnaC dependent ϕ X174 SSARF assay (dnaB protein = 7.8×10^5 ; Arai *et al.*, 1981b: dnaC = 6.0×10^5 ; Kobori and Kornberg, 1982b). From the quantities of dnaB protein (65 ng) and dnaC protein (42 ng) required for half-maximal activity, it was calculated that efficient replication of *ssiA*(R6K γ 2) *in vitro* required the use of excess dnaB protein (~14 molecules per DNA circle as hexamers, or ~84 protomers) and dnaC protein (~86 molecules per DNA circle as monomers). This represents a 4-fold increase over levels of these proteins required for efficient conversion of ϕ X174 single-stranded DNA to its replicative form (Kobori and Kornberg, 1982c).

Purification of dnaB and dnaC

A major consideration during the purification of the dnaB and dnaC proteins was the stability of the dnaB protein. It has been reported to be particularly unstable in the absence of ATP (Günther *et al.*, 1981; Arai *et al.*, 1981b). Addition of monovalent cations, ammonium sulfate, Mg^{2+} and glycerol each significantly increases the stability of the dnaB protein (Arai *et al.*, 1981b) and buffers used in the purification of the dnaB protein contained a combination of these additives at neutral pH. One further substantial benefit arising from the inclusion of ATP in these buffers is that it also satisfies the requirement for the maintenance of a dnaB-dnaC protein complex (Wahle *et al.*, 1989a), enhancing the possibility of isolating the intact complex from a cell-free extract.

Several procedures for lysis were examined with the thermoinduced strains RSC680 and RSC661. Although most protocols were capable of solubilizing the bulk of the overproduced proteins, the heat-shock lysozyme lysis procedure (Arai *et al.*, 1981b) was the most efficient and repeatable. Following this treatment, overproduced dnaB and dnaC proteins were extracted almost quantitatively into a soluble fraction as visualized by SDS-PAGE (Figure 5. 3). Assessment of crude extracts from these trials indicated that RSC680 produced ~1.5-fold more dnaB and dnaC protein than RSC661 (Table 5.2). RSC680 was therefore chosen as the source for purification of the dnaB and dnaC proteins. From a 5-l induced culture of this strain (7.17 g of cell paste), $\sim 16 \times 10^7$ units of each protein were extracted.

The dnaB and dnaC proteins in the soluble lysate both precipitated upon the addition of solid ammonium sulfate. Experimental trials revealed that the two overproduced proteins commenced precipitation with 0.16 g ammonium sulfate added per ml of

Table 5.2

Comparison of the relative activities of the dnaB and dnaC protein in Fraction I (Section 5.2.2) prepared from thermoinduced cultures of strains RSC661 and RSC680. Cultures (50 ml) were grown at 30 °C to $A_{595} \sim 0.5$. Cells were harvested from a portion (1 ml), and the remainder of the culture was shaken at 42 °C. After 4 h a further 1-ml sample was removed, then cells were harvested from 35 ml of culture, resuspended to $A_{595} = 25$ in lysis buffer and treated essentially as described in Section 5.2.2 for the large-scale lysis. The supernatant following centrifugation was assayed independently for dnaB and dnaC activity and for protein. The estimated quantity of each protein species was calculated as described (Figure 3.12) using the specific activity determined for the highly-purified dnaB and dnaC proteins (Table 5.3). From these results, strain RSC680 (containing the plasmid bearing the dnaB-dnaC synthetic operon) was calculated to have levels of overproduction of the dnaB and dnaC proteins ~ 1.5 -fold higher than levels determined for strain RSC661 (containing the pPS436 derivative).

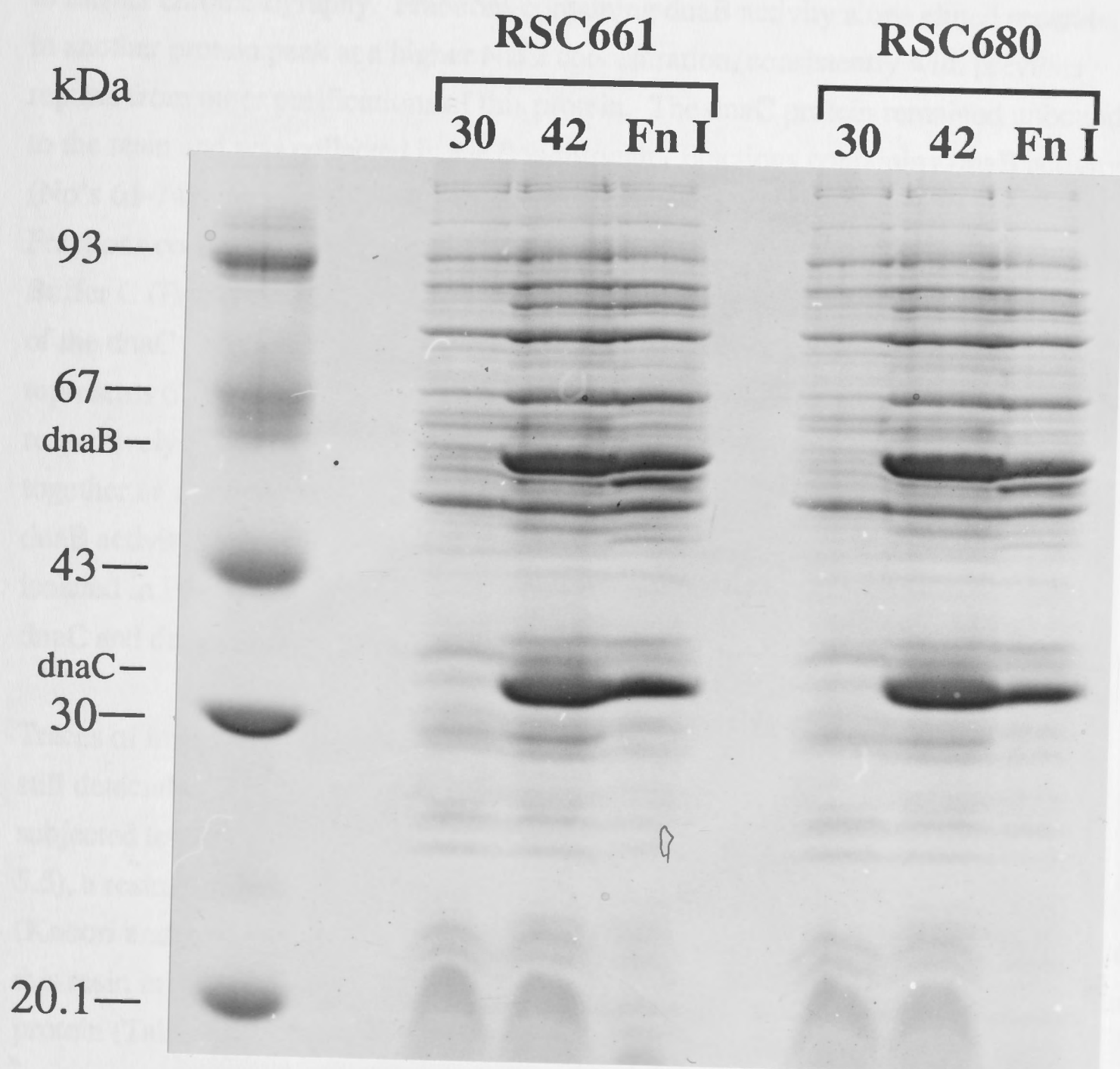
Figure 5.3

SDS-polyacrylamide gel electrophoresis of cells and Fractions I. Cells from the 1-ml portions of cultures RSC661 and RSC680 (see Table 5.2) removed prior to (30) and after 4 h of treatment at 42 °C (42) were resuspended to $A_{595} = 10$ in an SDS-gel loading buffer and heated for 2 min at 95 °C. Portions (20 μ l) were loaded onto a 12-% polyacrylamide gel, along with similarly treated samples (1000 units of dnaB protein) of Fraction I (Frn I). Protein markers were as described (Section 2.4.3). The proteins labelled dnaB (52.3 kDa) and dnaC (27.9 kDa) overproduced at 42 °C comigrated with authentic samples of these proteins (not shown).

Table 5.2

Source	Protein Assay	$10^{-6} \times$ Activity	$10^{-3} \times$ Specific Activity	Molecules/cell
		<i>units</i>	<i>units/mg</i>	
RSC661 (pPS554)	dnaB	3.20	20.3	216,000
	dnaC	3.48	22.1	253,000
RSC680 (pPS562)	dnaB	5.07	40.2	343,000
	dnaC	5.00	39.3	360,000

Figure 5.3



Fraction I (27.5 % saturation), although the isolation of all soluble dnaB and dnaC protein, as a pellet following centrifugation, was only possible after the addition of 0.2 g solid ammonium sulfate per ml (34 % saturation). This step resulted in isolation of more than 99% of the activity of each protein accompanied by >4-fold enrichment (Table 5.3), and therefore contributed significantly to the removal of the bulk of soluble contaminants from the preparation (Figure 5.7). The ammonium sulfate pellet containing dnaB and dnaC proteins was dissolved in a minimum volume of buffer containing factors required for stability of the dnaB protein. The residual ammonium sulfate was removed by dialysis and the solution was diluted to 100 ml (Fraction II), effectively reducing the conductivity and protein concentration of the sample to allow efficient binding to DEAE-cellulose resin.

The first protein peak to elute from the DEAE-cellulose column following application of the NaCl gradient was coincident with both peaks of dnaB and dnaC activities (Figure 5.4A). Samples visualized on SDS-PAGE revealed that protein in the peak fraction was comprised almost entirely of dnaB and dnaC protein (Figure 5.4B).

These fractions (No's 45-56; Fraction IIIB/C) were pooled and stored at -70 °C prior to further chromatography. Fractions containing dnaB activity alone eluted separately in another protein peak at a higher NaCl concentration, consistently with previous reports from other purifications of this protein. The dnaC protein remained unbound to the resin and was collected in the flowthrough. Fractions containing dnaB activity (No's 61-74; Fraction IIIB) were pooled and stored at -70 °C until required.

Fractions containing dnaC activity (No's 8-27) were also pooled and dialysed against Buffer C (Fraction IIIC) before immediate application to the next stage in purification of the dnaC protein. The total yields of dnaB and dnaC activity in these two fractions represents 67 % and 61 % of the total activity of each of these proteins in Fraction II respectively (Table 5.3). Virtually all of the remainder of the two proteins eluted together as the dnaB-dnaC complex (Fraction IIIB/C), *i.e.* 35 % of dnaC and 31 % of dnaB activity in Fraction II. When taken into consideration with those activities isolated in Fractions IIIC and IIIB, this represents a 93 % and 96 % recovery of all dnaC and dnaB activity, respectively, from the ammonium sulfate fraction.

Traces of impurities with the dnaC protein isolated in the flowthrough fraction were still detectable by SDS-PAGE (Figure 5.7). For this reason, Fraction IIIC was subjected to one further chromatography step using blue dextran-Sepharose (Figure 5.5), a resin with which substantial enrichment of dnaC had previously been reported (Kobori and Kornberg, 1982b). Purification of dnaC protein from Fraction IIIC with this resin in greater than 95 % yield was accompanied by a 1.09-fold enrichment of the protein (Table 5.2). The dnaC activity was coincident with a single peak of protein

Figure 5.4A

Separation on DEAE-cellulose of dnaB and dnaC proteins and the dnaB-dnaC complex from RSC680 (pPS562) cells. Conditions for purification are as described in Section 5.2.2. The resuspended and dialysed ammonium sulfate pellet was diluted to 100 ml with Buffer A (Fraction II) and applied to a column (16.3 x 2.5 cm) of DEAE-cellulose anion-exchange resin (DE-52) pre-equilibrated in Buffer A. The column was washed with Buffer A (105 ml) and bound proteins were eluted with a linear gradient (800 ml) of 0.025 to 0.6 M NaCl in Buffer A at a flow rate of 45 ml h. Fractions (7.5 ml each) were collected and assayed for dnaB-dnaC activity (I), dnaB activity alone (II), dnaC activity alone (III) and for protein (IV). For each protein species, those fractions pooled are indicated by brackets above the peak.

A.

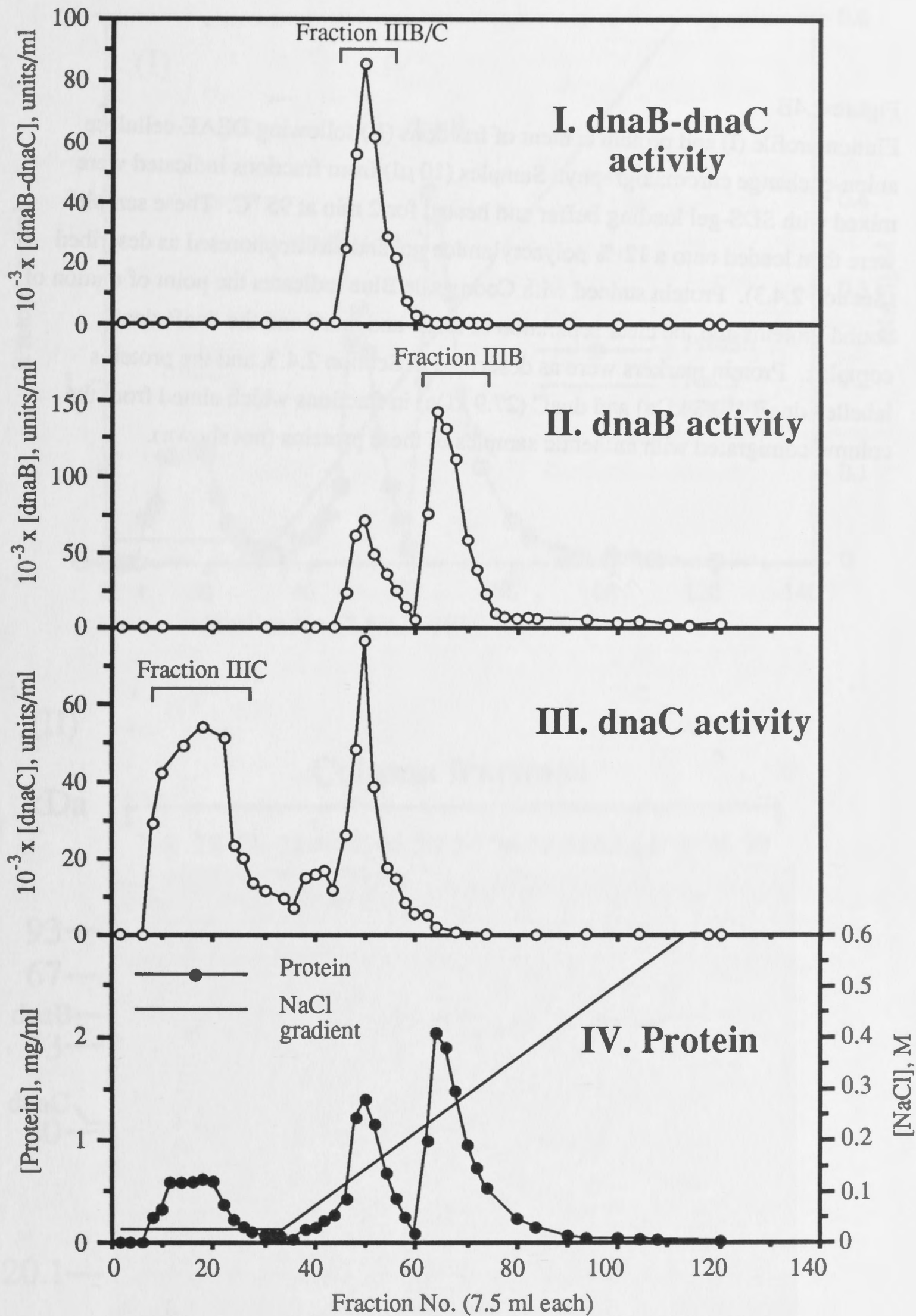
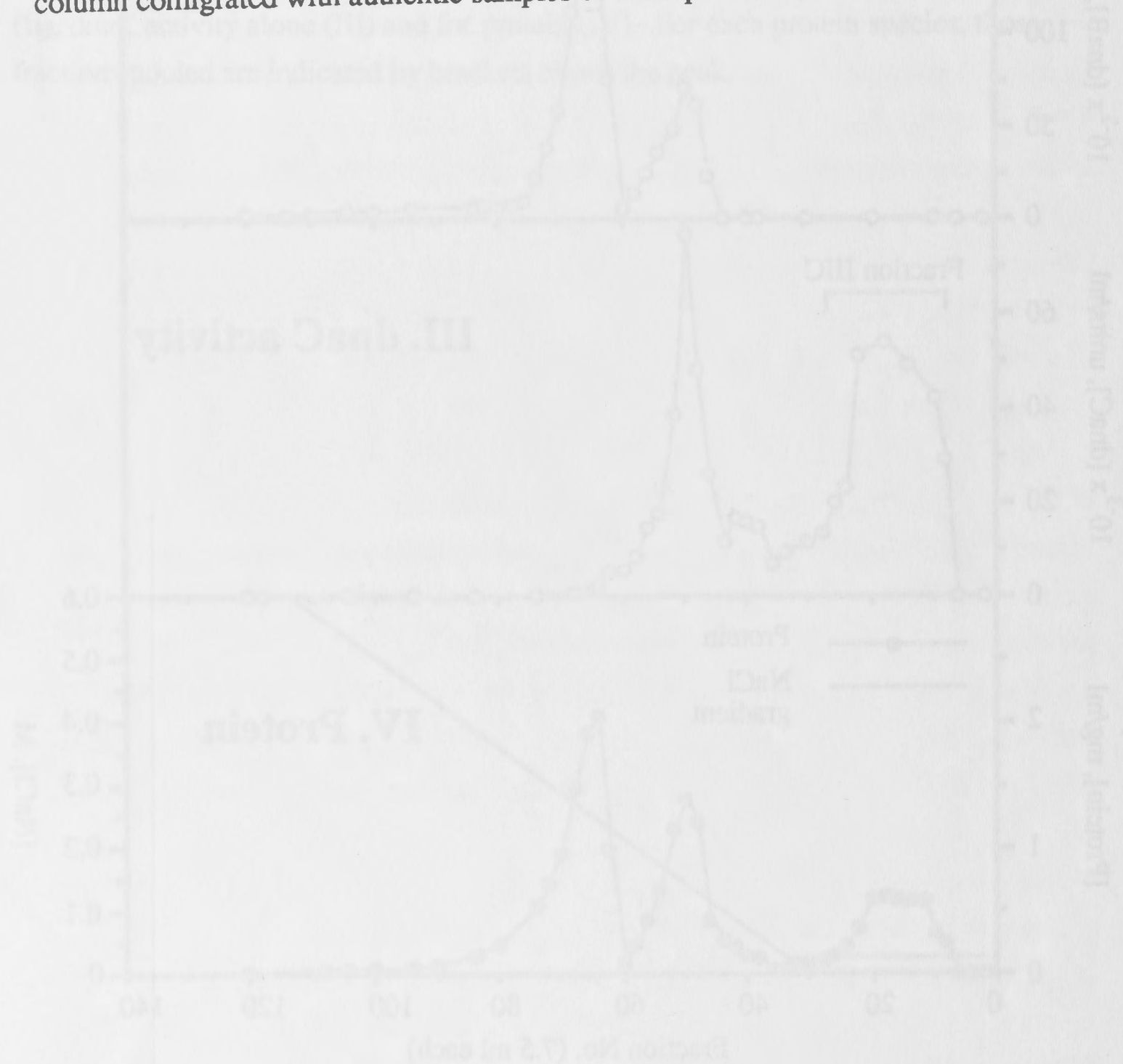
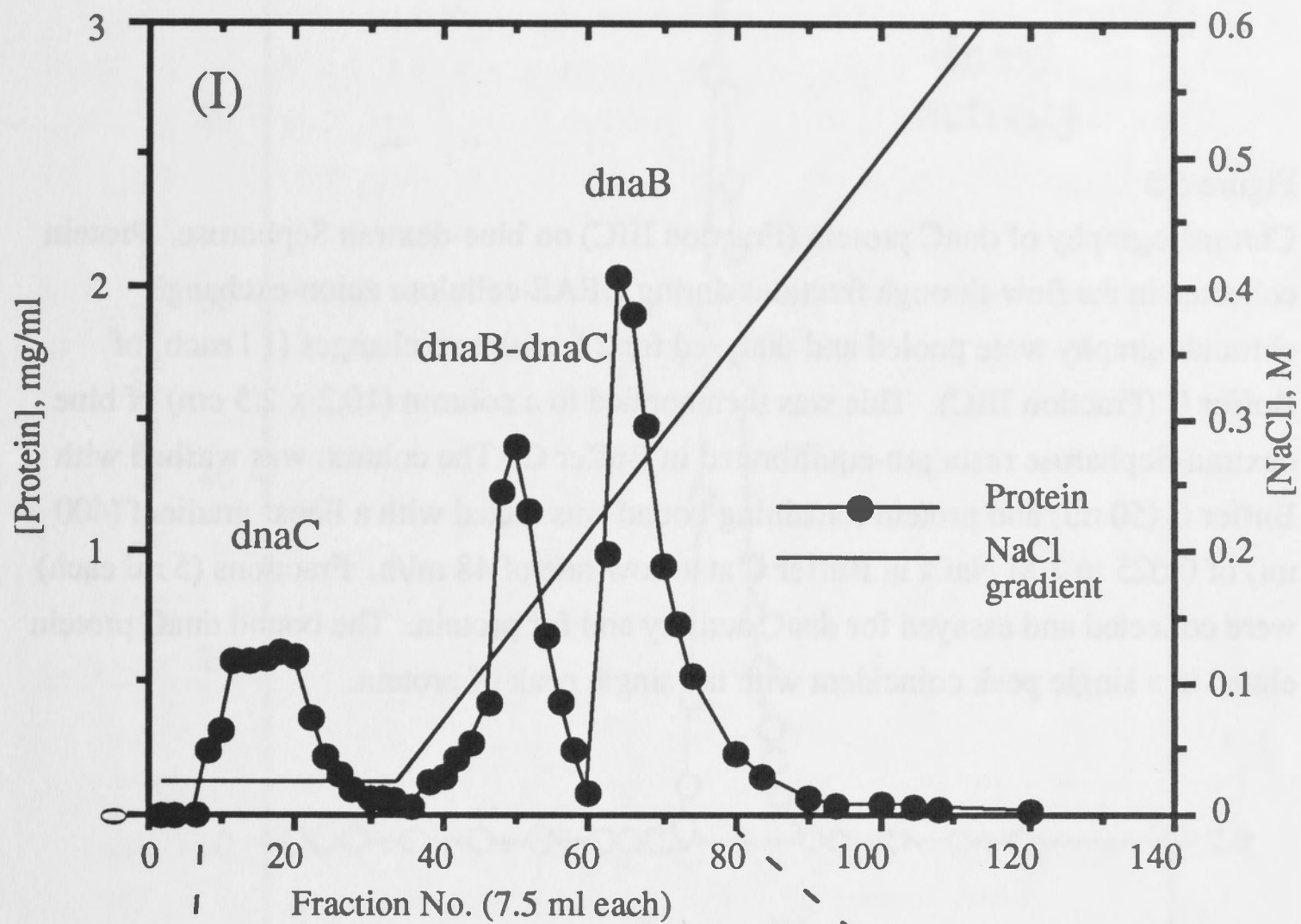


Figure 5.4B

Elution profile (I) and protein content of fractions (II) following DEAE-cellulose anion-exchange chromatography. Samples (10 μ l) from fractions indicated were mixed with SDS-gel loading buffer and heated for 2 min at 95 °C. These samples were then loaded onto a 12-% polyacrylamide gel and electrophoresed as described (Section 2.4.3). Protein stained with Coomassie Blue indicates the point of elution of bound proteins and the clear separation of dnaB and dnaC and the dnaB-dnaC complex. Protein markers were as described in Section 2.4.3, and the proteins labelled dnaB (52.3 kDa) and dnaC (27.9 kDa) in fractions which eluted from the column comigrated with authentic samples of these proteins (not shown).



B.



(II)

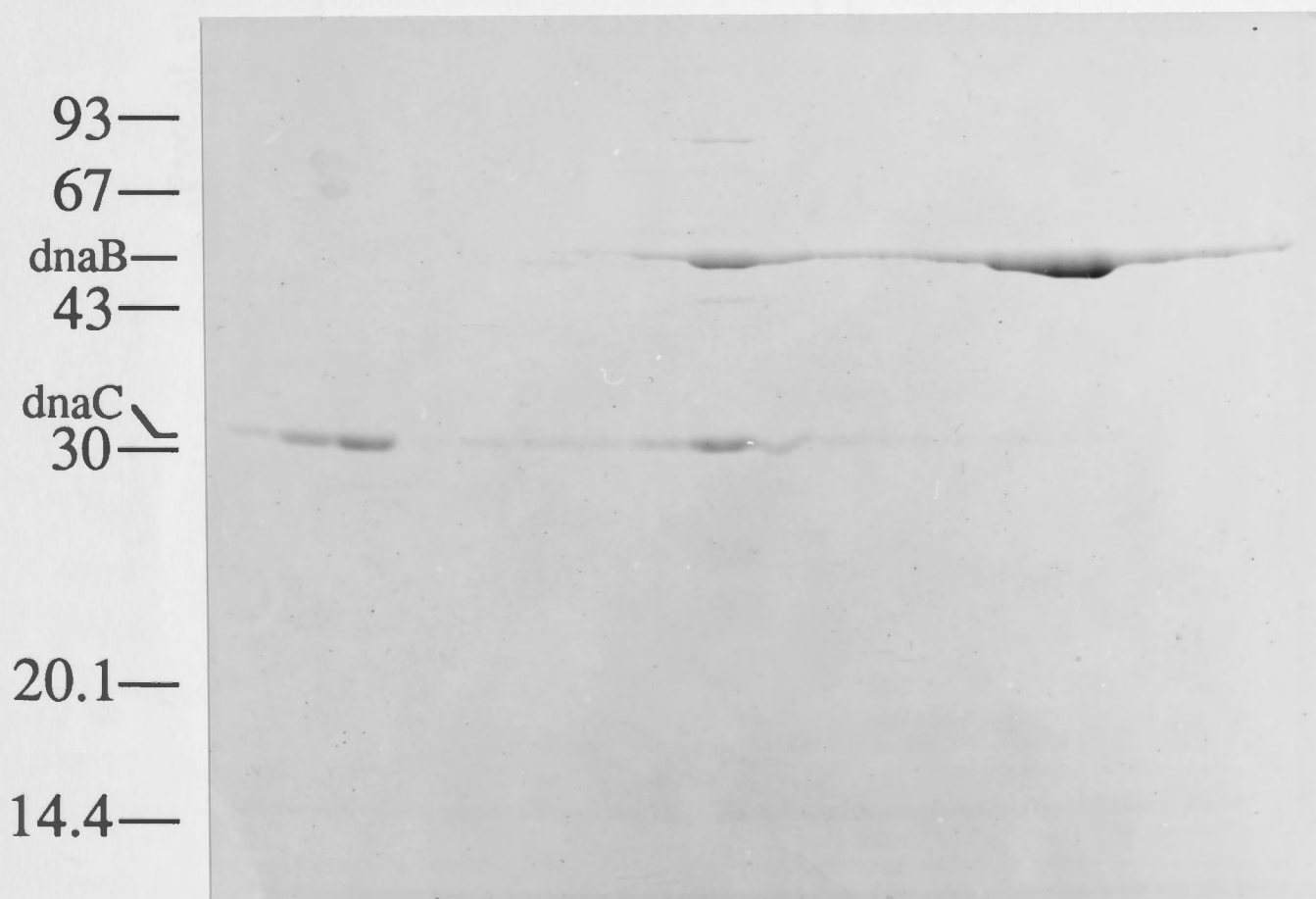
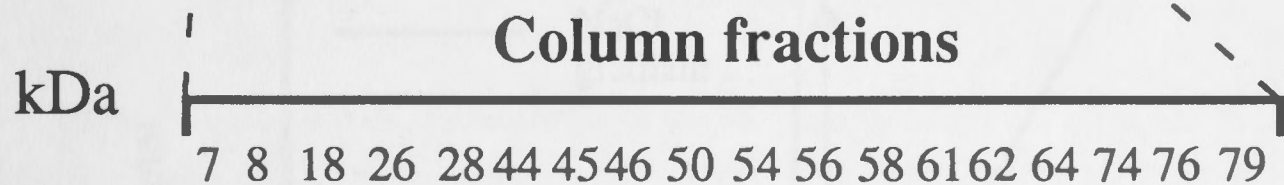
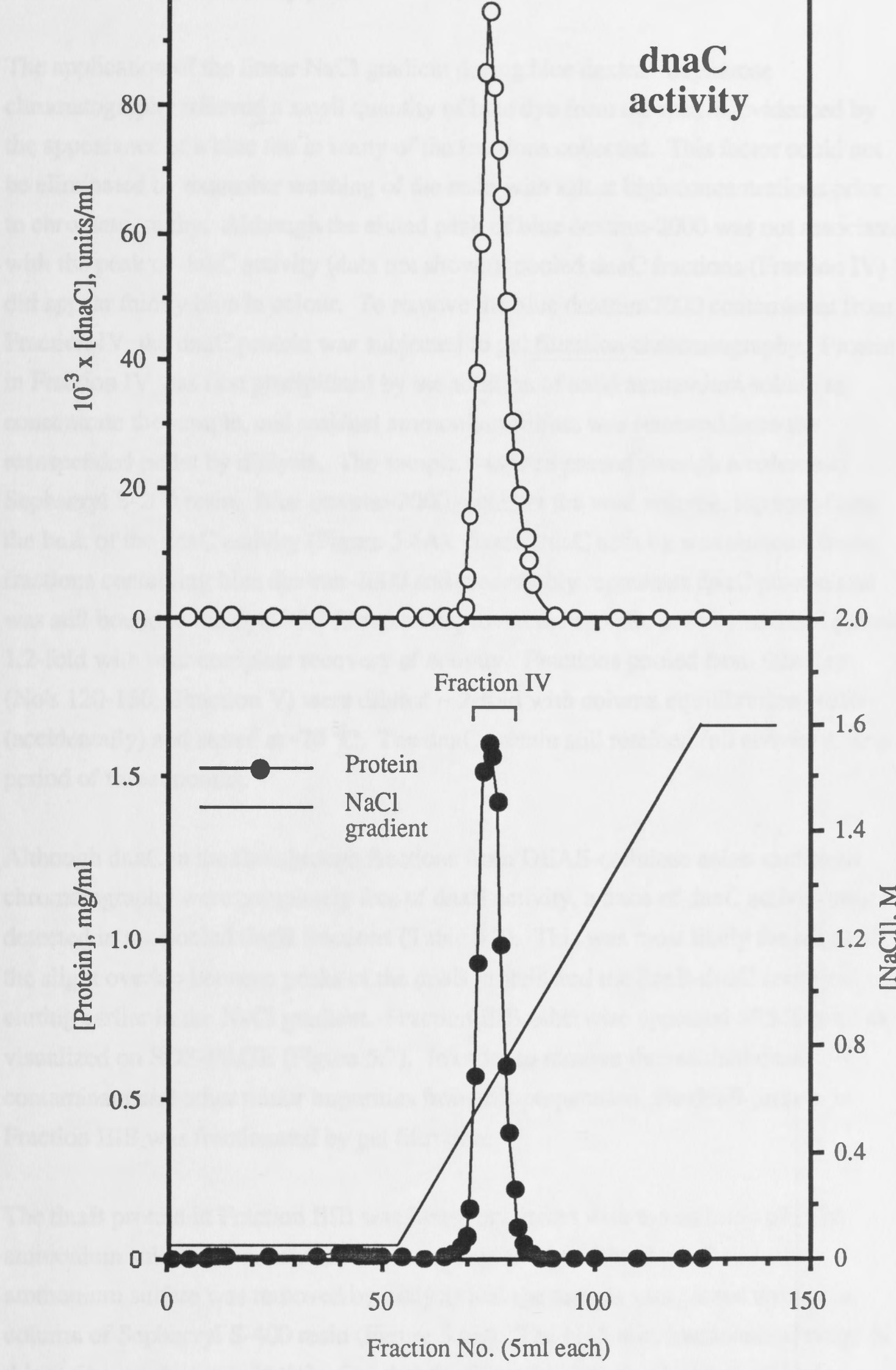


Figure 5.5

Chromatography of dnaC protein (Fraction IIIC) on blue-dextran Sepharose. Protein collected in the flow-through fractions during DEAE-cellulose anion-exchange chromatography were pooled and dialysed for 2 h with two changes (1 l each) of Buffer C (Fraction IIIC). This was then applied to a column (10.2 x 2.5 cm) of blue dextran-Sepharose resin pre-equilibrated in Buffer C. The column was washed with Buffer C (50 ml) and protein remaining bound was eluted with a linear gradient (400 ml) of 0.025 to 2 M NaCl in Buffer C at a flow rate of 48 ml/h. Fractions (5 ml each) were collected and assayed for dnaC activity and for protein. The bound dnaC protein eluted in a single peak coincident with the single peak of protein.

Figure 3. The effect of NaCl concentration on the activity of the *dnaC* protein purified following this study (Table 1). The *dnaC* protein was purified from the *Salmonella* strain 1402/79. The protein was purified by the method of [10]. The protein was purified by the method of [10]. The protein was purified by the method of [10].



(Figure 5.7), suggesting that dnaC protein pooled following this step (No's 71 to 81; Fraction IV) was essentially pure.

The application of the linear NaCl gradient during blue dextran-Sepharose chromatography relieved a small quantity of blue dye from the matrix, evidenced by the appearance of a blue tint in many of the fractions collected. This factor could not be eliminated by extensive washing of the resin with salt at high concentrations prior to chromatography. Although the eluted peak of blue dextran-2000 was not associated with the peak of dnaC activity (data not shown), pooled dnaC fractions (Fraction IV) did appear faintly blue in colour. To remove the blue dextran-2000 contaminant from Fraction IV, the dnaC protein was subjected to gel filtration chromatography. Protein in Fraction IV was first precipitated by the addition of solid ammonium sulfate to concentrate the sample, and residual ammonium sulfate was removed from the resuspended pellet by dialysis. The sample was then passed through a column of Sephacryl S-200 resin. Blue dextran-2000 eluted at the void volume, separate from the bulk of the dnaC activity (Figure 5.6A). Some dnaC activity was detected in the fractions containing blue dextran-2000 and presumably represents dnaC protein that was still bound to the dye. Gel filtration improved the specific activity of dnaC protein 1.2-fold with near complete recovery of activity. Fractions pooled from this step (No's 120-150; Fraction V) were diluted ~ 2-fold with column equilibration buffer (accidentally) and stored at -70 °C. The dnaC protein still retained full activity after a period of three months.

Although dnaC in the flowthrough fractions from DEAE-cellulose anion exchange chromatography were completely free of dnaB activity, a trace of dnaC activity was detected in the pooled dnaB fractions (Table 5.3). This was most likely the result of the slight overlap between peaks of the dnaB protein and the dnaB-dnaC complex eluting earlier in the NaCl gradient. Fraction IIIB otherwise appeared >95 % pure as visualized on SDS-PAGE (Figure 5.7). In order to remove the residual dnaC contaminant and other minor impurities from this preparation, the dnaB protein in Fraction IIIB was fractionated by gel filtration.

The dnaB protein in Fraction IIIB was first precipitated with the addition of solid ammonium sulfate to concentrate the sample prior to gel filtration. Residual ammonium sulfate was removed by dialysis and the sample was passed through a column of Sephacryl S-400 resin (Figure 5.6B). The high size fractionation range of this resin was chosen in anticipation that dnaB would elute as a large multimeric protein. Indeed, dnaB activity eluted early in the column profile. During this procedure the dnaB protein was enriched 1.08-fold with greater than 97 % recovery of

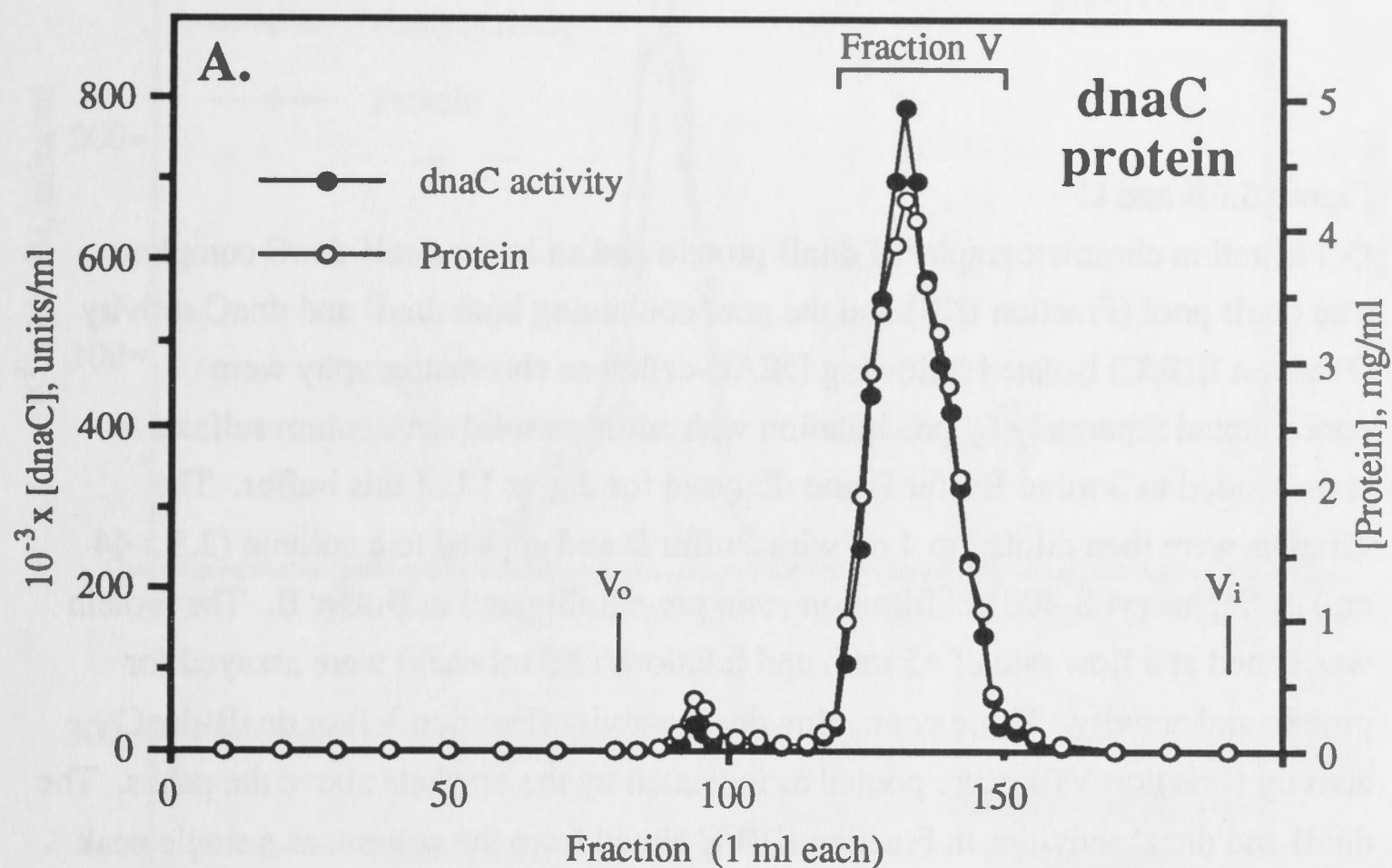
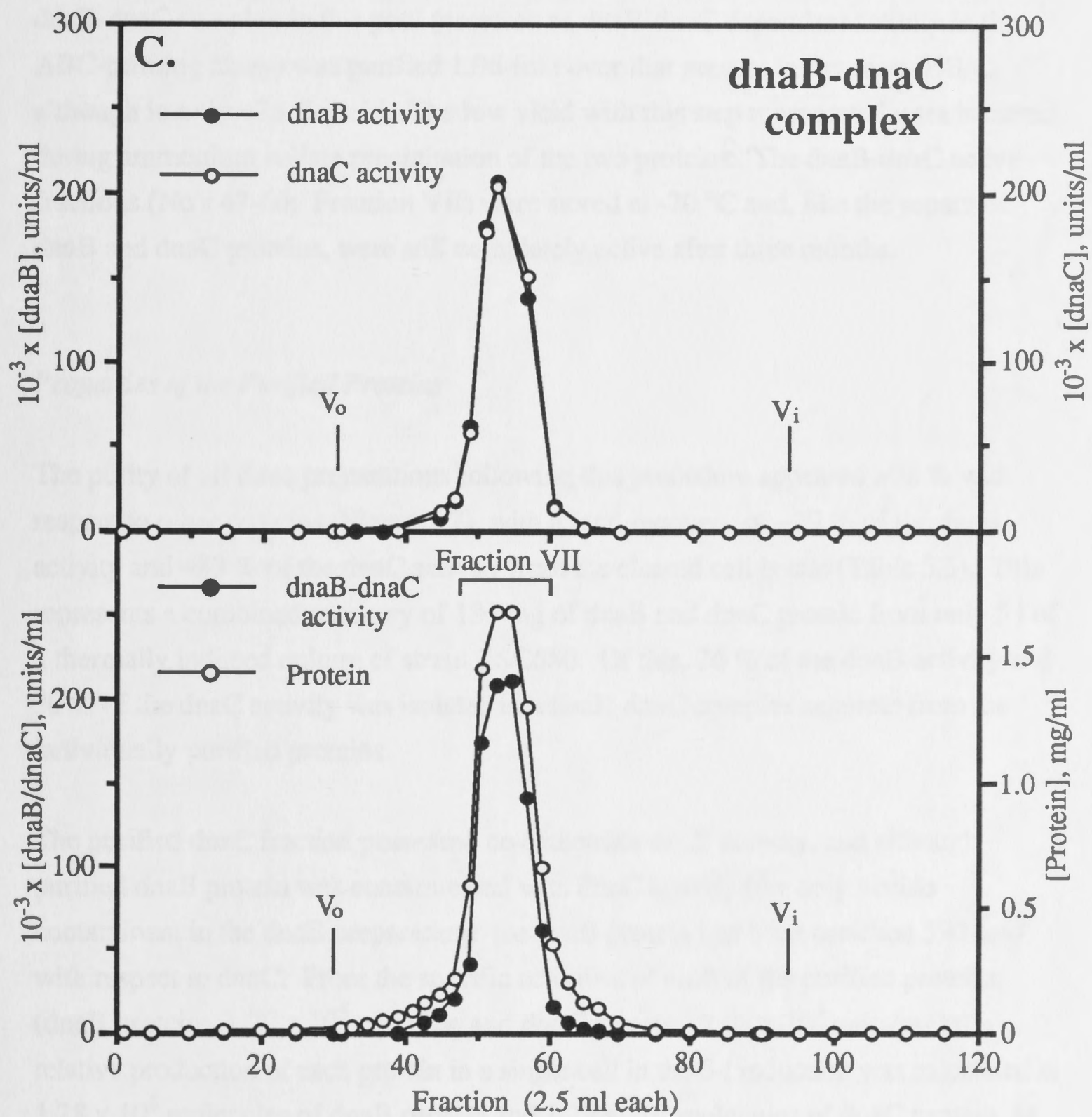
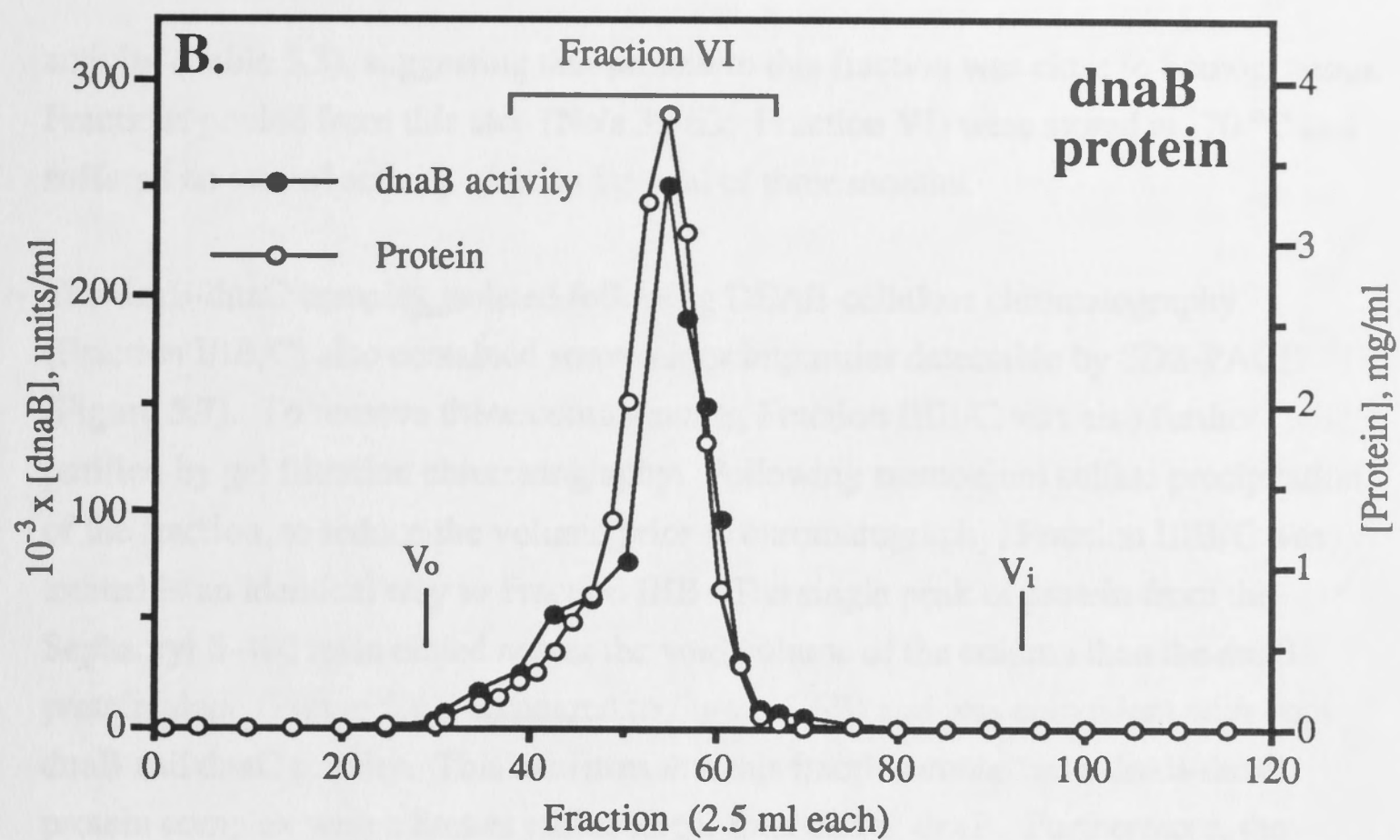


Figure 5.6A

Gel filtration chromatography of dnaC protein. The dnaC protein isolated following blue dextran-Sepharose chromatography (Fraction IV) was concentrated by precipitation with addition solid ammonium sulfate, resuspended to 3 ml in Buffer D and dialysed for 2 h vs 1 l of this buffer. The sample was then diluted to 4 ml with Buffer D and applied to a column (2.5 x 44 cm) of Sephacryl S-200 gel filtration resin pre-equilibrated in Buffer D. The protein was eluted at a flow rate of 45 ml/h; fractions (1 ml each) were collected and assayed for dnaC activity and for protein. Some dnaC protein eluted at the void volume of the column (V_o) and was probably associated with the residual blue dextran-2000 present in Fraction IV. The fractions pooled from this step are indicated by the bracket above the peak and V_o and V_i were determined as described (Section 2.4.5).

Figure 5.6B and C

Gel filtration chromatography of dnaB protein and an intact dnaB-dnaC complex. The dnaB pool (Fraction IIIB) and the pool containing both dnaB and dnaC activity (Fraction IIIB/C) isolated following DEAE-cellulose chromatography were concentrated separately by precipitation with addition solid ammonium sulfate, resuspended to 3 ml in Buffer B and dialysed for 2 h vs 1 l of this buffer. The samples were then diluted to 4 ml with Buffer B and applied to a column (2.5 x 44 cm) of Sephacryl S-400 gel filtration resin pre-equilibrated in Buffer B. The protein was eluted at a flow rate of 45 ml/h and fractions (2.5 ml each) were assayed for protein and activity. Those containing dnaB activity (Fraction VI) or dnaB-dnaC activity (Fraction VII), were pooled as indicated by the brackets above the peaks. The dnaB and dnaC activities in Fraction IIIB/C eluted from the column as a single peak coincident with a single peak of protein and suggest that these two proteins are intimately associated as a complex. The peak of dnaB-dnaC activity that eluted from this column eluted earlier than the dnaB protein alone in the column profile consistent with the larger Stokes radius expected for a dnaB-dnaC complex. The values for V_o and V_i were determined as described (Section 2.4.5).



activity (Table 5.3), suggesting that protein in this fraction was close to homogeneous. Fractions pooled from this step (No's 39-63; Fraction VI) were stored at -70°C and suffered no loss of activity after an interval of three months.

The dnaB-dnaC complex isolated following DEAE-cellulose chromatography (Fraction IIIB/C) also contained some minor impurities detectable by SDS-PAGE (Figure 5.7). To remove these contaminants, Fraction IIIB/C was also further purified by gel filtration chromatography. Following ammonium sulfate precipitation of the fraction, to reduce the volume prior to chromatography, Fraction IIIB/C was treated in an identical way to Fraction IIIB. The single peak of protein from the Sephacryl S-400 resin eluted nearer the void volume of the column than the dnaB protein alone (Figure 5.6C compared to Figure 5.6B) and was coincident with both dnaB and dnaC activity. This confirms that this fraction contained a dnaB-dnaC protein complex with a Stokes radius larger than that of dnaB. Furthermore, the dnaB-dnaC complex in this pool (assessed as dnaB-dnaC dependent activity in the ABC-priming assay) was purified 1.06-fold over that present in Fraction IIIB/C, although in only ~75 % yield. The low yield with this step represents losses incurred during ammonium sulfate precipitation of the two proteins. The dnaB-dnaC active fractions (No's 47-60; Fraction VII) were stored at -70°C and, like the separated dnaB and dnaC proteins, were still completely active after three months.

Properties of the Purified Proteins

The purity of all three preparations following this procedure appeared $>98\%$ with respect to other proteins (Figure 5.7), with a total recovery of ~89 % of the dnaB activity and ~83 % of the dnaC activity from the cleared cell lysate (Table 5.3). This represents a combined recovery of 181 mg of dnaB and dnaC protein from only 5 l of a thermally induced culture of strain RSC680. Of this, 26 % of the dnaB activity and 30 % of the dnaC activity was isolated in a dnaB-dnaC complex separate from the individually purified proteins.

The purified dnaC fraction possessed no detectable dnaB activity, and although purified dnaB protein was contaminated with dnaC activity (the only visible contaminant in the dnaB preparation), the dnaB protein had been enriched 590-fold with respect to dnaC. From the specific activities of each of the purified proteins (dnaB protein; 1.20×10^5 units/mg and dnaC protein; 2.08×10^5 units/mg) the relative production of each protein in a single cell in the 5-l induction was calculated at 1.28×10^5 molecules of dnaB protein and 1.29×10^5 molecules of dnaC protein, as

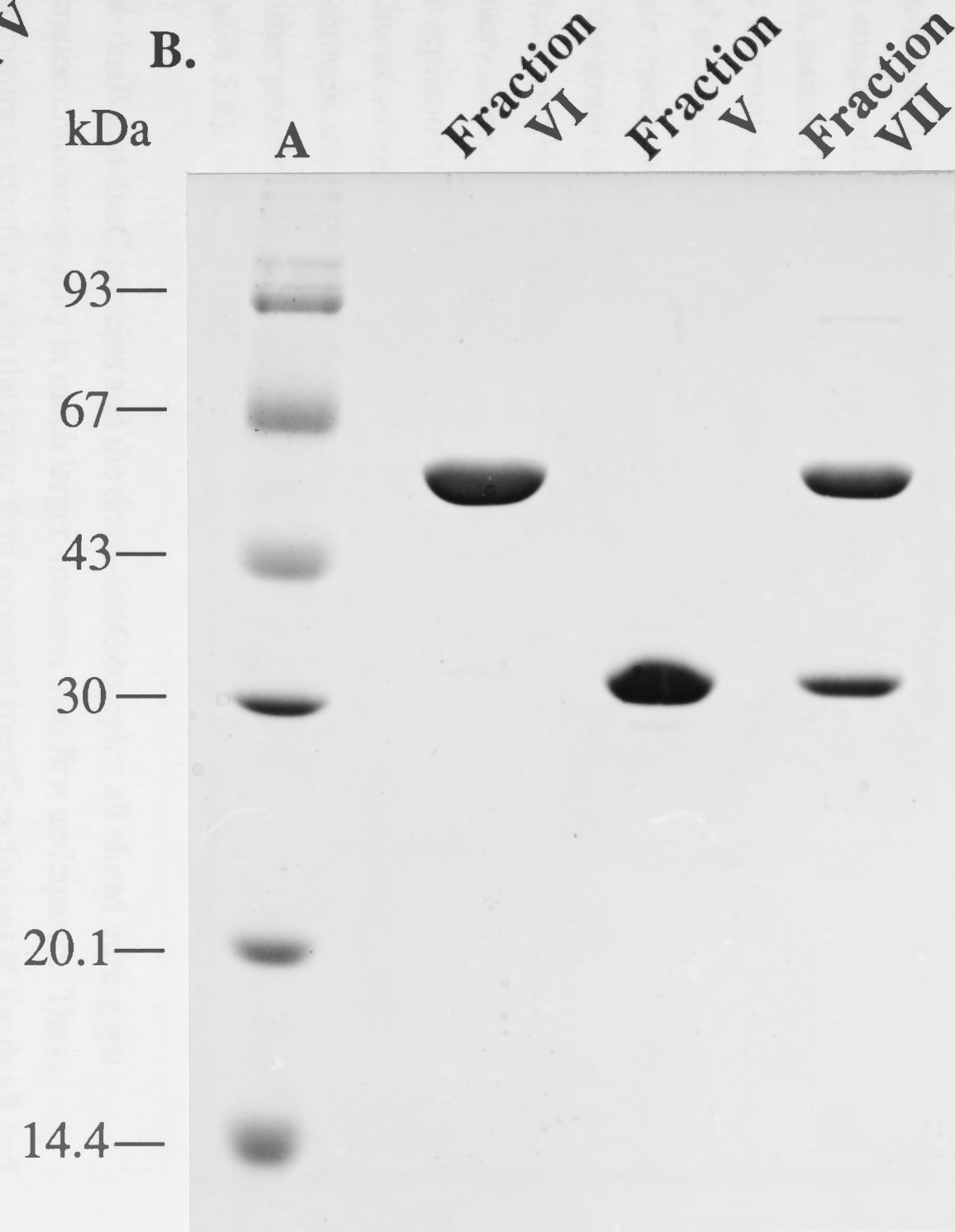
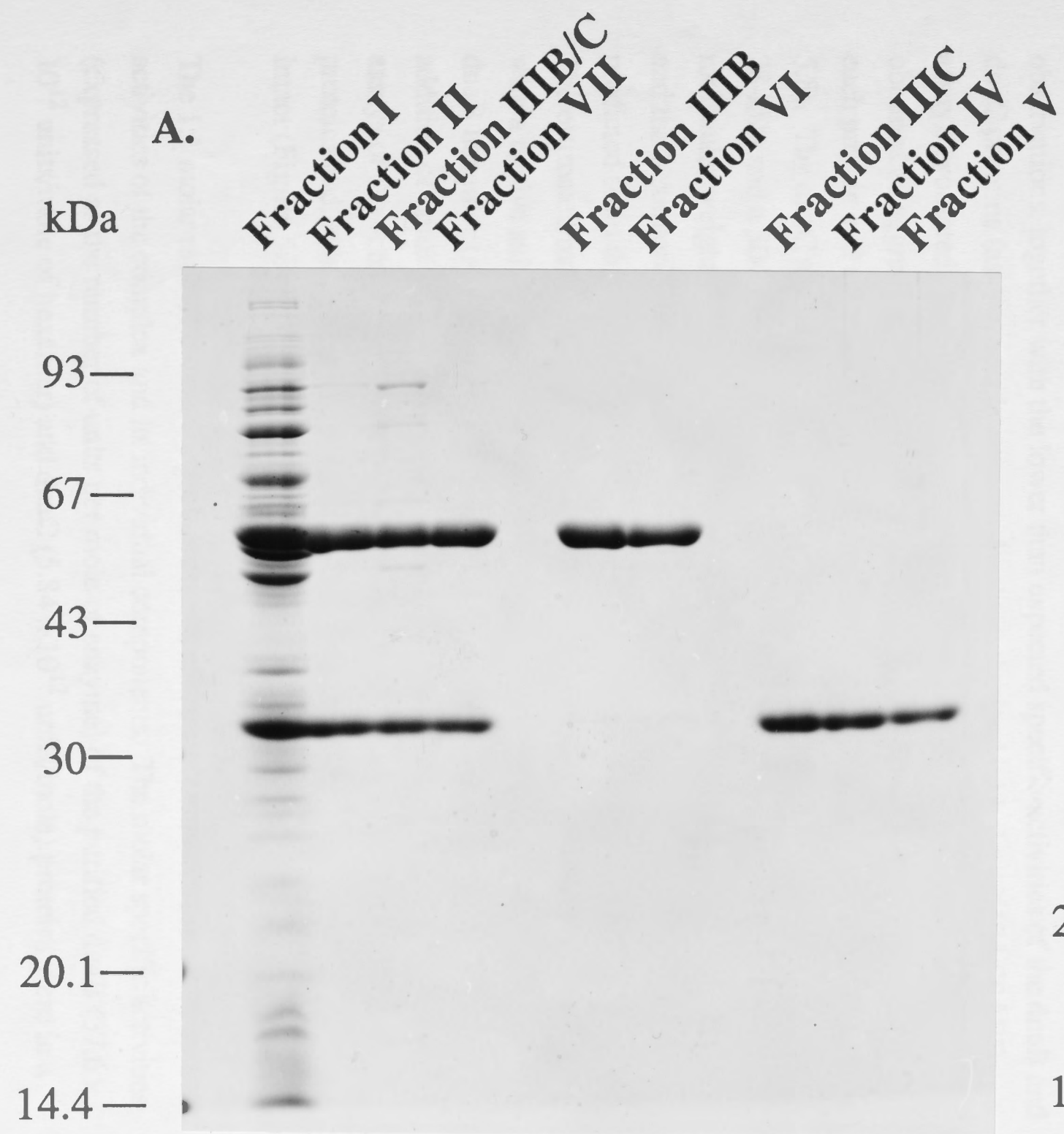
Table 5.3

Purification of dnaB protein, dnaC protein and dnaB-dnaC complex from strain RSC680. Cells from 5 l of cell culture cells were lysed using a heat-shock lysozyme lysis and the overproduced protein products extracted almost quantitatively into the cleared cell lysate (Fraction I). Following precipitation with ammonium sulfate (Fraction II), the dnaC and dnaB proteins and a dnaB-dnaC complex were separated by DEAE-cellulose chromatography (Fractions IIIC, B and B/C). The dnaC protein was further purified by blue dextran-Sepharose chromatography (Fraction IV), and all three preparations were fractionated using gel filtration. Fractions I - VII were assayed for dnaB, dnaC and dnaB-dnaC activity using the ABC-priming assay (Section 5.2.1) and for protein. Activities below the limits of detection are indicated by dashed lines. Values for each step in the purification take into account the 1-ml samples removed for analysis of the previous fraction.

Fraction	Protein	dnaC				dnaB				dnaB-dnaC complex			
		$10^{-6} \times$ Activity	$10^{-3} \times$ Specific Activity	Yield	Purification	$10^{-6} \times$ Activity	$10^{-3} \times$ Specific Activity	Yield	Purification	$10^{-6} \times$ Activity	$10^{-3} \times$ Specific Activity	Yield	Purification
	<i>mg</i>	<i>units</i>	<i>units/mg</i>	<i>%</i>	<i>-fold</i>	<i>units</i>	<i>units/mg</i>	<i>%</i>	<i>-fold</i>	<i>units</i>	<i>units/mg</i>	<i>%</i>	<i>-fold</i>
I. Lysis supernatant	1171	15.9	13.6	(100)	-----	16.9	14.4	(100)	-----	17.3	14.8	(100)	-----
II. Ammonium sulfate	280	15.7	56.1	99.1	4.1	16.9	60.4	100	4.2	17.1	61.1	99.2	4.1
IIIC. DEAE-cellulose dnaC pool	58.9	9.5	161	61.1	11.9	<0.01	-----	-----	-----	<0.01	-----	-----	-----
IV. Blue dextran-Sephadex pool	51.4	9.0	175	57.8	13.0	<0.01	-----	-----	-----	<0.01	-----	-----	-----
V. Sephadex S-200 pool	42.8	8.9	208	58.3	15.5	<0.01	-----	-----	-----	<0.01	-----	-----	-----
IIIB. DEAE-cellulose dnaB pool	101	0.05	0.5	-----	-----	11.2	111	66.9	7.7	0.06	0.6	-----	-----
VI. Sephadex S-400 dnaB pool	90.3	0.02	0.2	-----	-----	10.8	120	65.2	8.3	0.03	0.3	-----	-----
IIIB/C. DEAE-cellulose dnaB-dnaC pool	69.3	5.1	73.6	32.5	5.4	5.1	73.6	30.1	5.1	5.2	75.0	30.1	5.0
VII. Sephadex S-400 dnaB-dnaC pool	48.3	3.8	79.1	24.6	5.8	3.8	78.9	22.8	5.5	3.8	79.1	22.4	5.3

Figure 5.7

Purification of dnaB protein, dnaC protein and dnaB-dnaC complex from strain RSC680. A. Proteins in Fractions I-VII were separated by 12.5-% polyacrylamide gel electrophoresis in the presence of SDS. Lanes 1-4: Fraction I, Fraction II, Fraction IIIB/C and Fraction IV (1,000 units dnaB-dnaC complex); lanes 5 and 6: Fraction IIIB and Fraction VI (1,000 units of dnaB protein); lanes 6-8, Fraction IIIC, Fraction IV and Fraction V (1,000 units of dnaC protein). Protein markers (sizes in kDa) were as in Section 2.4.3. B. Highly-purified dnaB protein (Fraction VI, 30 μ g), dnaC protein (Fraction V, 30 μ g) and dnaB-dnaC complex (Fraction VII, 30 μ g) were purified as described (Section 5.2.2; Table 5.3) and separated by 12.5-% polyacrylamide gel electrophoresis in the presence of SDS. Protein markers in lane A (sizes in kDa) were as in Section 2.4.3.



monomers (using the formula outlined in Chapter 3; footnote #5). These figures may be compared to the levels of the proteins overproduced in smaller (<50 ml) cultures. Calculations based on the activity of dnaB and dnaC protein in a Fraction I from a 50-ml thermoinduced culture of RSC680 (Figure 5.3) indicated that 3.4×10^5 and 3.6×10^5 molecules of dnaB and dnaC respectively had been produced after 4 h at 42 °C. This represents a ~3-fold increase over levels detected in a similarly treated 5-l culture of the strain and suggests that dnaB and dnaC protein comprise >50 % of the total soluble protein of cells in these small-scale inductions. This result is consistent with observations with other strains directing the overexpression of other genes involved in the replication of the *E. coli* chromosome in that lower level overproduction in larger cultures seems to occur quite often (Chapter 3; Dr N.E. Dixon, personal communication). Metal ion analysis of the pure dnaB and dnaC proteins revealed that neither protein was associated with a significant quantity of those metal ions analysed (Table 5.4).

The dnaB and dnaC proteins and the dnaB-dnaC complex all eluted during gel filtration chromatography in much larger volumes than first anticipated. These observations, together with the lower than expected specific activities of the dnaB and dnaC proteins (as compared to reported values obtained with the ϕ X174 SSARF assay), prompted concern as to whether the proteins purified from RSC680 were obtained as aggregated species. For this reason, the apparent molecular weight of each protein and the dnaB-dnaC complex was determined by gel filtration (Figure 5.8). The dnaC and dnaB proteins eluted as sharp peaks at positions expected for a 28-kDa and a 304-kDa protein respectively. These values are in accordance with the molecular weights of monomeric dnaC protein ($M_r = 27,900$; Masai and Arai, 1988b) and the hexameric dnaB protein ($M_r = 6 \times 52,300$; Nakayama *et al.*, 1984b) as predicted from the calculated size of polypeptides translated from each gene sequence. The estimated molecular weight of the dnaB-dnaC complex was 470 kDa, consistent with a twelve subunit complex comprised of six monomers of dnaC protein and a dnaB hexamer (Kobori and Kornberg, 1982c; Wahle *et al.*, 1989a). Neither the addition of dnaB or dnaC protein to the dnaB-dnaC complex stimulated its activity in assays of DNA replication suggesting that the complex formed between these two proteins and isolated from a thermoinduced culture of RSC680 is also completely intact (Figure 5.9).

The 1:1 molar ratio estimated by gel filtration is also consistent with the molar specific activities of the complex and its individual components. The molar specific activities (expressed as the number of units per mole of enzyme) of the purified dnaB (37.6×10^{12} units/mole of hexamer) and dnaC (5.84×10^{12} units/mole) proteins were in a

Table 5.4.

The metal ion content of highly-purified dnaB and dnaC. Samples (10 ml) from preparations described in Section 5.2.2 (Fractions VI and VII, respectively) were dialysed as outlined (Section 2.4.6). Specific activities were determined as described in Section 5.2.1. Protein concentrations were determined spectrophotometrically, using molecular weights of 52,300 (dnaB protein) and 27,900 (dnaC protein), and $\epsilon_{280} = 29,870 \text{ M}^{-1} \text{ cm}^{-1}$ and $23,470 \text{ M}^{-1} \text{ cm}^{-1}$ (estimated from the amino acid compositions; Gill and von Hippel, 1989) for pure dnaB and dnaC, respectively. Metal ions were determined by atomic absorption spectrometry as described (Section 2.4.6) using flame atomization. The dnaB protein contained 0.02 g atom/mol of Zn and 0.04 g atom of Cu, whereas the dnaC protein contained 0.02 g atom/mol of Zn and 0.04 g atom of Cu. Analysis of both these samples for other metal ions (Mn, Fe, Co and Ni) gave levels below the limits of detection (<0.02 g atom/mol of dnaB or dnaC). The loss of specific activity during dialysis of the dnaB and dnaC samples (58 % and 32 %, respectively, compared to that of the native proteins; Table 5.3) is probably not related to loss of intrinsic metal ion but rather to the removal of stabilizing buffer components.

Sample	$10^{-3} \times$ Specific Activity	[Protein]	Metal Ion	[Metal Ion]
	<i>units/mg</i>	μM		μM
dnaB (Fraction VI)	49.9	17.2	Zn Cu	0.31 0.63
dnaC (Fraction VII)	141.5	14.9	Zn Cu	0.61 0.63

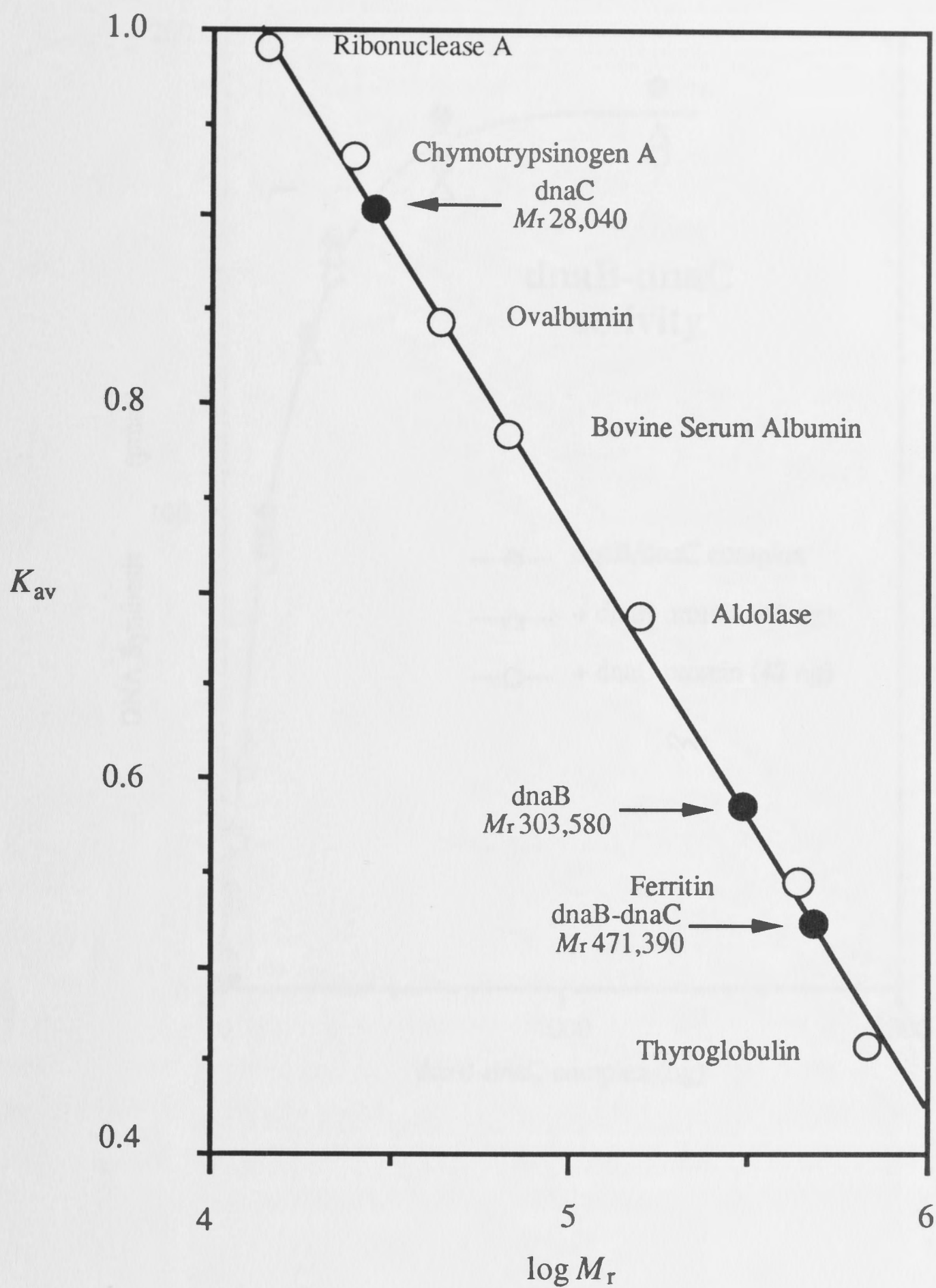


Figure 5.8

Molecular size of dnaB, dnaC and the dnaB-dnaC complex estimated by gel filtration. Proteins were chromatographed on Sephacryl S-400 as described in Section 2.4.5. The K_{av} values $[(V_e - V_o)/(V_i - V_o)]$ determined for each of the purified proteins, the dnaB-dnaC complex and the standards (Section 2.4.5), are plotted. The line was determined by linear regression analysis.

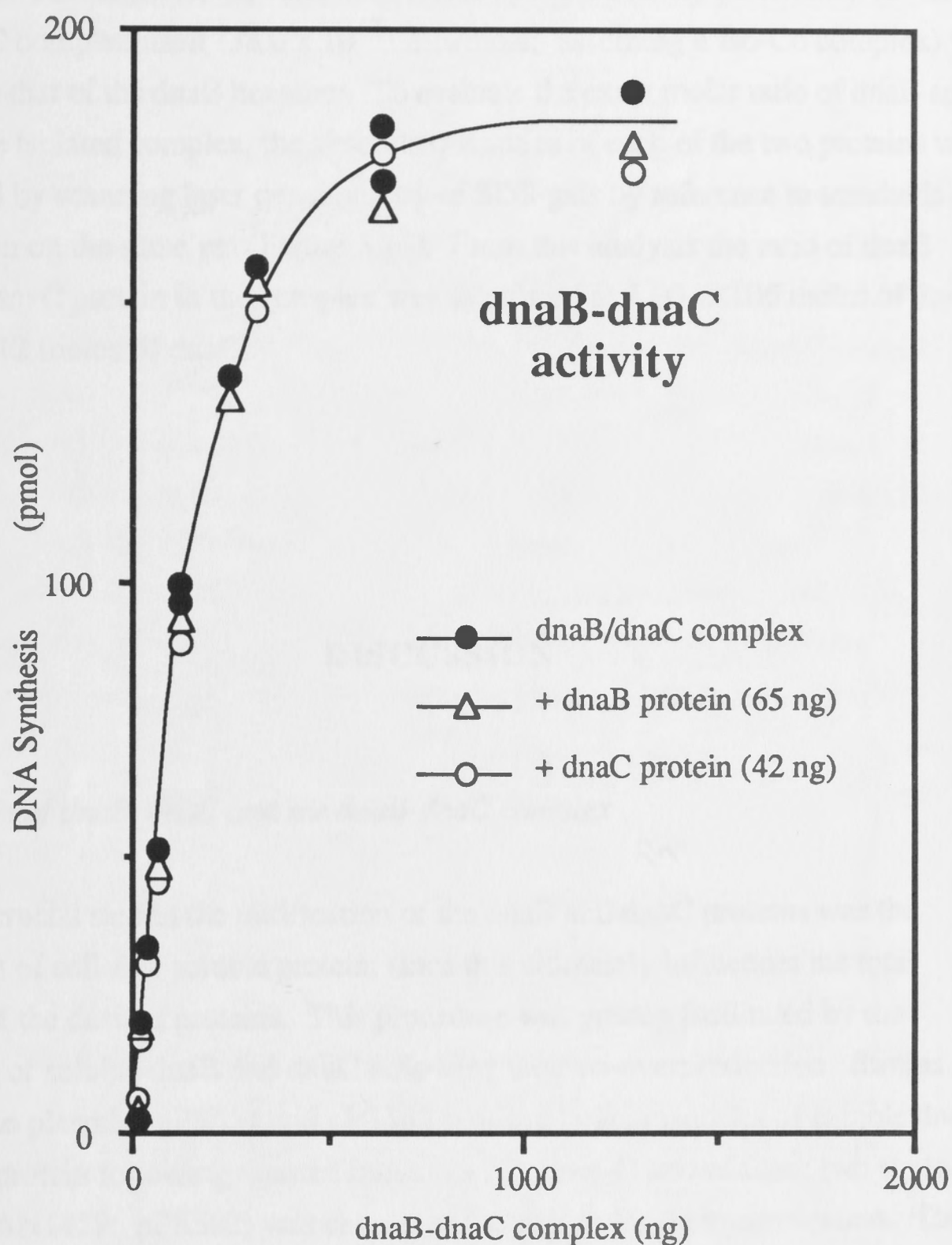


Figure 5.9

The dnaB-dnaC activity of Fraction VII prepared from a thermally-induced strain (RSC680) and containing the plasmid pPS562 (Section 5.5.2). The addition of highly-purified dnaB protein (Fraction VI; 65 ng) and dnaC protein (Fraction V; 42 ng) in quantities calculated to give half-maximal activity of these proteins in their respective assays had no effect on the level of activity of the dnaB-dnaC complex.

6:0.93 ratio. Furthermore, the calculated molar specific activity determined for the dnaB-dnaC complex itself (38.0×10^{12} units/mole; assuming a B6-C6 complex) was identical to that of the dnaB hexamer. To evaluate the exact molar ratio of dnaB and dnaC in the isolated complex, the absolute quantities of each of the two proteins was determined by scanning laser densitometry of SDS-gels by reference to standards of each protein on the same gel (Figure 5.10). From this analysis the ratio of dnaB protein to dnaC protein in the complex was calculated as 1.00 ± 0.06 moles of dnaB to 1.075 ± 0.12 moles of dnaC.

DISCUSSION

Purification of dnaB, dnaC and the dnaB-dnaC complex

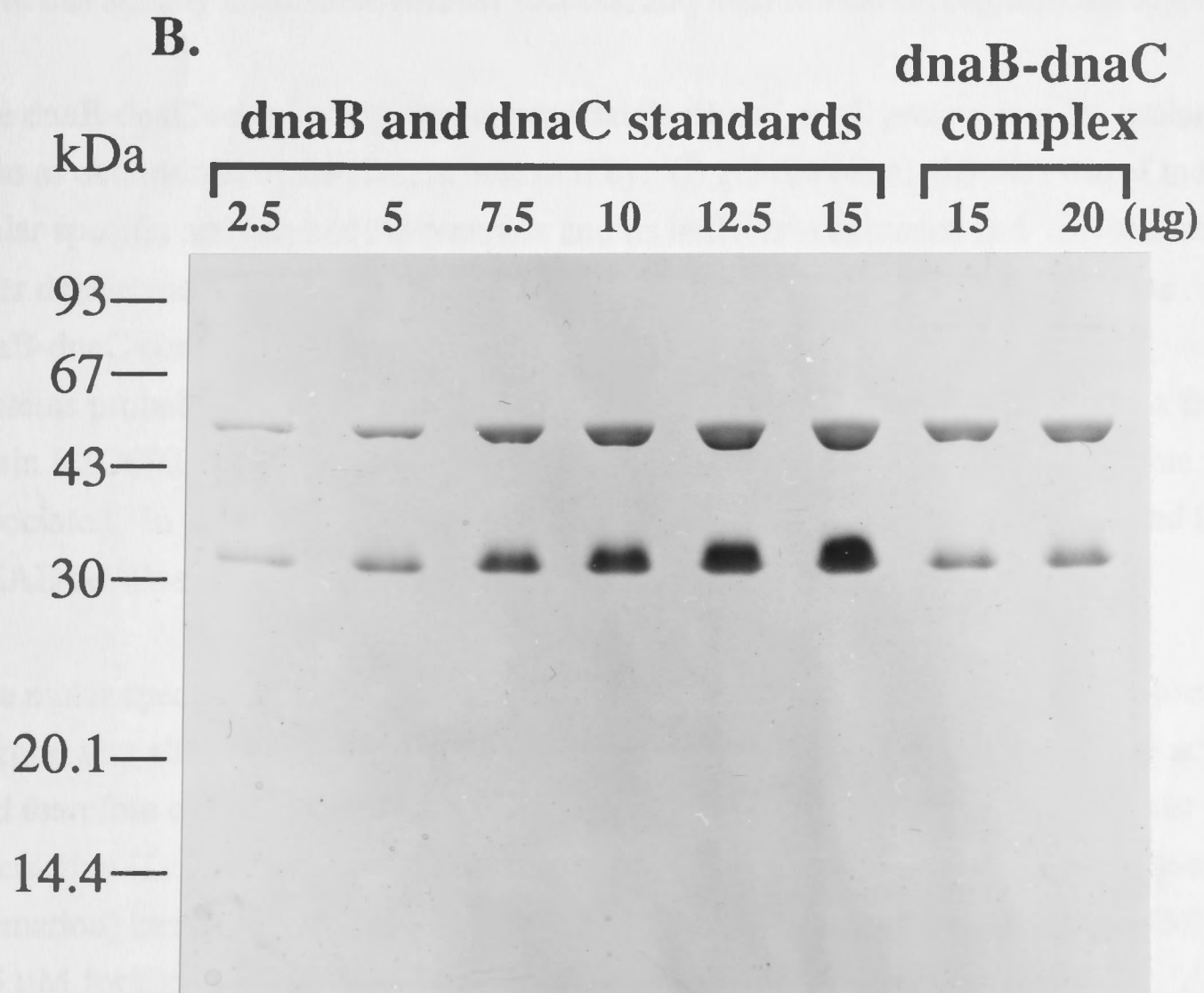
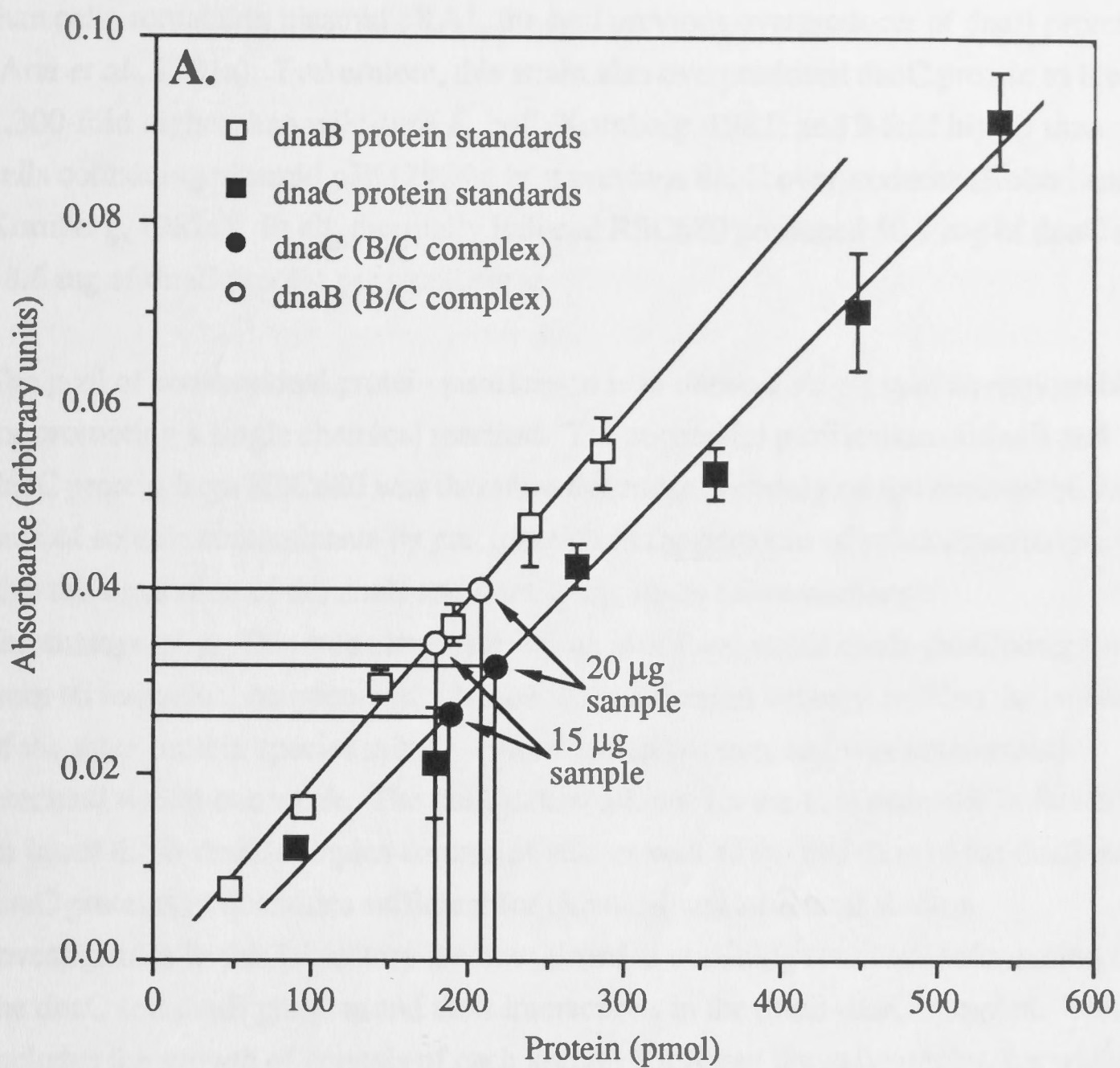
The most crucial step in the purification of the dnaB and dnaC proteins was the preparation of cell-free soluble protein, since this ultimately influences the total recovery of the desired proteins. This procedure was greatly facilitated by the generation of soluble dnaB and dnaC following their co-overproduction. Strains carrying the plasmids pPS554 and pPS562 produce large quantities of soluble dnaB and dnaC protein following thermal induction (Chapter 4) and of these two strains, RSC680 (AN1459; pPS562) was chosen as the source for their purification. The production of a single transcript encoding both *dnaC* and *dnaB* from this plasmid seemed to give near equimolar amounts of each protein, and they were produced more efficiently than in a strain carrying a plasmid with the *dnaB* and *dnaC* genes transcribed from separate promoters.

The total activity of dnaB and dnaC in a lysate of RSC680 was almost identical suggesting that the two proteins were extracted in almost equivalent quantities following thermal induction. The slight differences, within statistical error, may well represent translational efficiency associated with each gene and/or efficiency with which each protein is used in the ABC priming assay, since there was no evidence of any significant amount of either that remained in the insoluble fraction. From these values, and the specific activities of the pure protein, it was possible to estimate the post-induction cellular levels of each. Strain RSC680 overproduced dnaB to levels 6,400-fold higher than wild-type *E. coli* cells (Ueda *et al.*, 1978) and 32-fold higher

Figure 5.10

Estimation of the molar ratio of dnaB and dnaC protein in the dnaB-dnaC complex was obtained by reference to standards of each protein using scanning laser densitometry of SDS-gels. B. Samples of highly-purified dnaB and dnaC protein, prepared as described (Section 5.2.2, Table 5.3), were loaded onto a 12-% polyacrylamide gel in the presence of SDS. Equivalent amounts of dnaB and dnaC (ranging from 2.5 to 15 μg ; determined by the dye-binding method of Bradford, 1976) were loaded together in lanes denoted "dnaB and dnaC standards". Two samples (15 and 20 μg) of similarly quantified dnaB-dnaC complex were loaded on two other lanes of the same gel. After electrophoretic separation protein bands were stained with Coomassie blue (Section 2.4.3). Using the procedure described in Section 2.4.4, an estimation of the absorbance of a known quantity of protein in each band was determined. A. Plotting the results obtained for the highly-purified individual proteins enabled the construction of a standard curve for each species, using linear regression analysis to calculate the line of best fit and determination of standard errors. Those absorbance values obtained for the separated proteins from the dnaB-dnaC complex were then interpolated to gain the amount of each protein in the complex. From these data it was calculated that this preparation of dnaB-dnaC complex contained 1.00 ± 0.06 moles of dnaB to 1.075 ± 0.12 moles of dnaC.

Two notable sources of potential error in this analysis were: (i) the assumption that the Bradford assay gives a reliable measure of the amounts of dnaB and dnaC used as standards, and (ii) that accurate values for the quantity of protein in a band could be obtained by single passes of a broad-width laser rather than densitometric determination of the total area of each band. The Bradford assay was used in preference to direct measurements of A_{280} (*cf.* Table 5.4) because of uncertainty of the contribution of bound ATP (from the buffers used) to ϵ_{280} . This problem could be overcome by standardization with nucleotide-free samples of dnaB and dnaC proteins.



than cells containing plasmid pKA1, the best previous overproducer of dnaB protein (Arai *et al.*, 1981a). Furthermore, this strain also overproduced dnaC protein to levels 1,300-fold higher than wild-type *E. coli* (Kornberg, 1982) and 9-fold higher than cells containing plasmid pJK129, the best previous dnaC overproducer (Kobori and Kornberg, 1982a). In all, thermally induced RSC680 produced 10.7 mg of dnaC and 18.6 mg of dnaB protein per g cell paste.

The goal of conventional protein purification is to obtain a single species responsible for promoting a single chemical reaction. The successful purification of dnaB and dnaC protein from RSC680 was therefore dependant not only on the removal of the bulk of soluble contaminants by precipitation in the presence of ammonium sulfate but also the separation of the dnaB and dnaC proteins by anion-exchange chromatography. This step also separated an intact and stable dnaB-dnaC complex from its individual components. Overall the purification strategy enabled the isolation of the three protein species in >90 % at each step but one, and was comfortably executed within one week. The purification allows, for the first time, the isolation of an intact dnaB-dnaC complex formed *in vivo* as well as the isolation of the dnaB and dnaC proteins in quantities sufficient for chemical and structural studies.

Investigations in this laboratory are now aimed at obtaining structural information on the dnaC and dnaB proteins and their interactions in the dnaB-dnaC complex. This includes the growth of crystals of each species for X-ray crystallography, for which there has already been some limited success, and high resolution electron microscopy.

The dnaB-dnaC complex isolated comprised dnaB and dnaC protein in a 1:1 molar ratio as determined by its size, as assessed by: (i) gel filtration; (ii) the ratio of the molar specific activities of the complex and its individual subunits; and (iii) scanning laser densitometry of components separated by SDS-PAGE. The stability of the dnaB-dnaC complex suggests that this is the state in which the dnaB and dnaC proteins probably occur *in vivo* (Wahle *et al.*, 1989a). Isolation of this complex from strain RSC680, however, indicates that not all dnaB and dnaC overproduced was associated. In the preparation described, only ~30 % of the two proteins isolated from DEAE-cellulose chromatography were obtained as the complex (Table 5.3).

The molar specific activities for the individual proteins and that of the complex itself suggest that all dnaB and dnaC protein overproduced in RSC680 is biologically active and therefore capable of interaction in the presence of ATP. The apparent K_M for this interaction (*i.e.*, that concentration of nucleotide required for half-maximal complex formation) has been determined at 15 μM for ATP (Kobori and Kornberg, 1982c) and 0.5 μM for the nonhydrolysable ATP γ S (Wahle *et al.*, 1989a), well below the

concentration of ATP used in chromatography buffers (Section 5.2).

This raises something of a paradox. While interaction of the two proteins appears to be necessary for them to be obtained in soluble form (Chapter 4), and the complex appears to be stable to dissociation (at least in the presence of ATP), substantial amounts of the separate components are easily resolved chromatographically. Note that gel filtration of the complex under conditions described in Figures 5.6 and 5.8 results in no detectable dissociation of the dnaC protein.

One possible reason for dissociation of the complex prior to or during DEAE-cellulose chromatography could be interactions between the complex and DNA present in Fractions I and II. Binding of the dnaB-dnaC complex to single-stranded DNA results in subsequent release of the dnaC protein; *i.e.* the dnaB-dnaC complex dissociates (Wahle *et al.*, 1989b). Under the *in vitro* buffer conditions with the low temperatures used during purification, the equilibrium between complex formation and disassociation in the presence of DNA may favour the maintenance of a high proportion of uncomplexed dnaB and dnaC proteins. This proposal is supported by the fact that the complex, once free from DNA contaminants, is perfectly stable during gel filtration (Figure 5.6c). The proportion of free proteins relative to the complex isolated varied somewhat from preparation to preparation (data not shown), although fortunately not all protein was complexed as this enabled the isolation of the individual components.

The ABC-priming assay

The purified dnaB and dnaC proteins described here had specific activities lower than anticipated. The values for dnaC protein (2.08×10^5 units/mg) and dnaB protein (1.20×10^5 units/mg) were respectively ~3-fold and ~6-fold lower than values obtained in the ϕ X174 SS \rightarrow RF assay (Kobori and Kornberg, 1982b; Ueda *et al.*, 1978; Arai *et al.*, 1981b). The assay used for the evaluation of activities for proteins from this preparation, the ABC priming assay, can replace ϕ X174-type priming for efficient lagging strand synthesis and duplex unwinding (Masai *et al.*, 1990b), but the mechanism of priming on *ssiA*(R6K γ 2) template is more akin to that performed by the *oriC*-type primosome assembled at the origin of replication (see Chapter 2).

Replication of all *oriC* plasmids by a reconstituted enzyme system requires the prior interaction of dnaA protein with specific sequences in the DNA template (Bramhill and Kornberg, 1988a). The efficient conversion of *ssiA*(R6K γ 2) single-stranded DNA

template into a duplex replicative form was similarly observed in the presence of dnaA, dnaB and dnaC proteins upon the addition of primase, DNA polymerase III holoenzyme and ribo- and deoxyribonucleotide substrates to the SSB-coated template. When SSB was omitted, dnaA and dnaC were no longer required since general 'non-specific' priming by dnaB and primase operates on single-stranded DNA template in the absence of single-stranded binding protein (Arai and Kornberg, 1981d, e). General priming was stimulated by the addition of dnaC (Table 5.1), as observed previously (Wahle *et al.*, 1989b).

Assembly of the *oriC*-type primosome is probably similar to assembly of the ϕ X-type primosome that occurs at *pas* sequences. Recognition of the template by the specific binding protein, dnaA or PriA, occurs first, and through subsequent protein-protein interactions the dnaB protein becomes stably established on the template. The primosome assembled on the ϕ X174 template, however, comprises all seven primosomal proteins and is thought to be similar to the primosome responsible for initiating discontinuous lagging strand DNA synthesis at the replication fork of *E. coli* chromosomal DNA. In ϕ X174 replication the PriA, PriB, PriC and dnaT proteins are required for loading of dnaB-dnaC complex onto the template. The dnaA protein presumably fulfils this function alone in ABC-priming.

The lower specific activities for the dnaB and dnaC proteins in the ABC assay as compared to the ϕ X174 assay is reflected in the four-fold difference calculated for the efficiency of dnaB and dnaC action. The consistency of values obtained for the specific activities of dnaB, dnaC and the complex (see above) argues that the lower values in the ABC assay are due to differences in the assay system, rather than the presence in the preparation of substantial amounts of inactive proteins. This will be tested by assay of the present protein in the ϕ X174 assay, when it is available to us. The reasons for the difference in efficiency of utilization of dnaB and dnaC in the two reactions may include some of the following: (i) Variations in the efficiency and stability of hairpin formation on each template; (ii) Differences between the affinities of dnaA and PriA protein for their respective recognition sites; (iii) Different mechanisms of primosome assembly, a process that requires the dnaC-dependent delivery of the dnaB protein to either the dnaA-DNA complex or the PriA-DNA complex. The latter complex may also contain the PriB, PriC and dnaT proteins which could increase the affinity of the dnaB-dnaC complex for protein PriA by forming a 'protein bridge' or scaffold between the two species; and (iv) Variation in the efficiency with which each primosome primes DNA replication. This may include the speed and efficiency of translocation on the single-stranded template and the nature of the functional interaction between primase and the dnaB protein.

CHAPTER 6

Involvement of a Leucine Heptad Repeat Motif in Oligomerization of dnaB Protein

In recent years three distinct structural motifs have been identified among DNA binding proteins. These are thought to be responsible for many of the protein-protein and protein-nucleic acid interactions involved in DNA metabolism and include the helix-turn-helix, zinc-finger and leucine zipper motifs (Struhl, 1989).

The leucine zipper motif was originally characterized on the basis of four or five leucine residues in a heptad repeat (*i.e.*, the leucine residues occur at every seventh amino acid in the polypeptide; Landschulz *et al.*, 1988). When modelled as an idealized α -helix, the leucyl residues are found to be aligned along one face and hydrophobic residues are distributed around them, forming an amphipathic helix. It was proposed that protein-protein binding was facilitated by the interaction of two such motifs to form a dimer of anti-parallel α -helices which are stabilized by the sequential interdigitation of their leucine residues to produce 'the molecular zipper'.

This structural motif is known to be responsible for the dimerization and subsequent target DNA recognition of several eukaryotic transcription factors including Fos, Jun and GCN4 (Turner and Tjian, 1989; Johnson and McKnight, 1989). Synthetic peptides corresponding to the leucine zippers of these proteins have now been constructed and circular dichroism spectroscopy has shown these to be α -helical structures (O'Shea *et al.*, 1989a, b).

These studies and more recent data obtained by X-ray crystallography (Rassmussen *et al.*, 1991) and mutational analysis (Hu *et al.*, 1990) have indicated that the molecular basis for the protein-protein interaction is not leucyl side chain interdigitation but the formation of a coiled-coil structure. This makes these heptad repeat sequences the shortest members of a class of proteins that form parallel helical coiled-coils. Evidence for this comes from the fact that the characteristic leucine repeat of these leucine zipper motifs occurs four residues out of phase with a second repeat of hydrophobic amino acids. These leucine side chains contribute to a '4-3 hydrophobic repeat' which is directly involved in the binding interface between the two polypeptides forming a parallel coiled-coil (Cohen and Parry, 1986). This has also been substantiated by two-dimensional NMR experiments confirming that the GCN4 leucine zipper homodimer is helical and forms a symmetrical dimer, as expected for a coiled coil (Oas *et al.*, 1990; Saudek *et al.*, 1990). However, unlike dimers formed by the hydrophobic interaction of two helices in a parallel coiled-coil, proteins identified as containing leucine zippers have an absolute requirement for the presence of a leucine heptad repeat (Hu *et al.*, 1990). A hydrophobic interface *per se* is not sufficient to encode a functional leucine zipper.

Scanning the primary amino-acid sequence of the dnaB protein, obtained from the nucleotide sequence of the gene (Nakayama *et al.*, 1984b), revealed a leucine heptad repeat toward the COOH-terminus. Alignment of the carboxy-terminal residues 361-392 of the *E. coli* dnaB protein with similar COOH-terminal sequences of dnaB-related proteins from other sources reveals a striking sequence similarity. Only 26 % of residues do not share identity between all five sequences and only one residue is different in more than two sequences. This four unit leucine heptad repeat motif commencing at isoleucine 361 (Figure 6.1) therefore possesses characteristics of known zipper motifs identified in many eukaryotic proteins. The repeating heptad suggests extensive potential α -helical structure, there is hydrophobic contribution from residues at position a and this region shares remarkable amino-acid sequence homology with the primary sequence of dnaB-like proteins from other sources.

The only aberration from the 'classic' leucine zipper model is the presence of a highly conserved proline residue at the centre of the motif. A computer search of a protein data base has revealed over 200 proteins that contain four or more leucines spaced seven residues apart (O'Shea *et al.*, 1989a). Many of these were discounted as potential oligomerization motifs because of proline residues found within the leucine repeats. Proline by its nature is a helix-breaking residue (Chou and Fasman, 1978) and has been shown to destabilize known leucine zipper dimers when substituted within the motif (Gentz *et al.*, 1989; Turner and Tjian, 1989; Hu *et al.*, 1990).

However, the dnaB protein from *E. coli* is known to associate as a hexameric structure (Arai *et al.*, 1981b) where the subunits in the dnaB multimer remain firmly associated and are not easily replaced. It is therefore unlikely that the protein-protein interactions in forming this stable complex are identical to those witnessed in eukaryotic transcriptional activators. It is quite possible that the proline at the centre of the dnaB leucine heptad repeat may impart structural identity necessary for dnaB protein oligomerization.

The amino-acid sequences of a great variety of proteins have been determined, but like the dnaB protein sequence they are usually derived from gene sequencing with the protein itself never structurally characterized. This has led to the widespread use of various algorithms applied to estimate local secondary structural elements within proteins (Chou and Fasman, 1978; Garnier *et al.*, 1978). Using these methods, secondary structural elements in the *E. coli* dnaB protein have been predicted (Nakayama *et al.*, 1984b; Figure 6.2). Based on these studies, the dnaB protein was considered to contain two structurally independent domains separated by a hinge region comprised extensively of α -helix.

A.

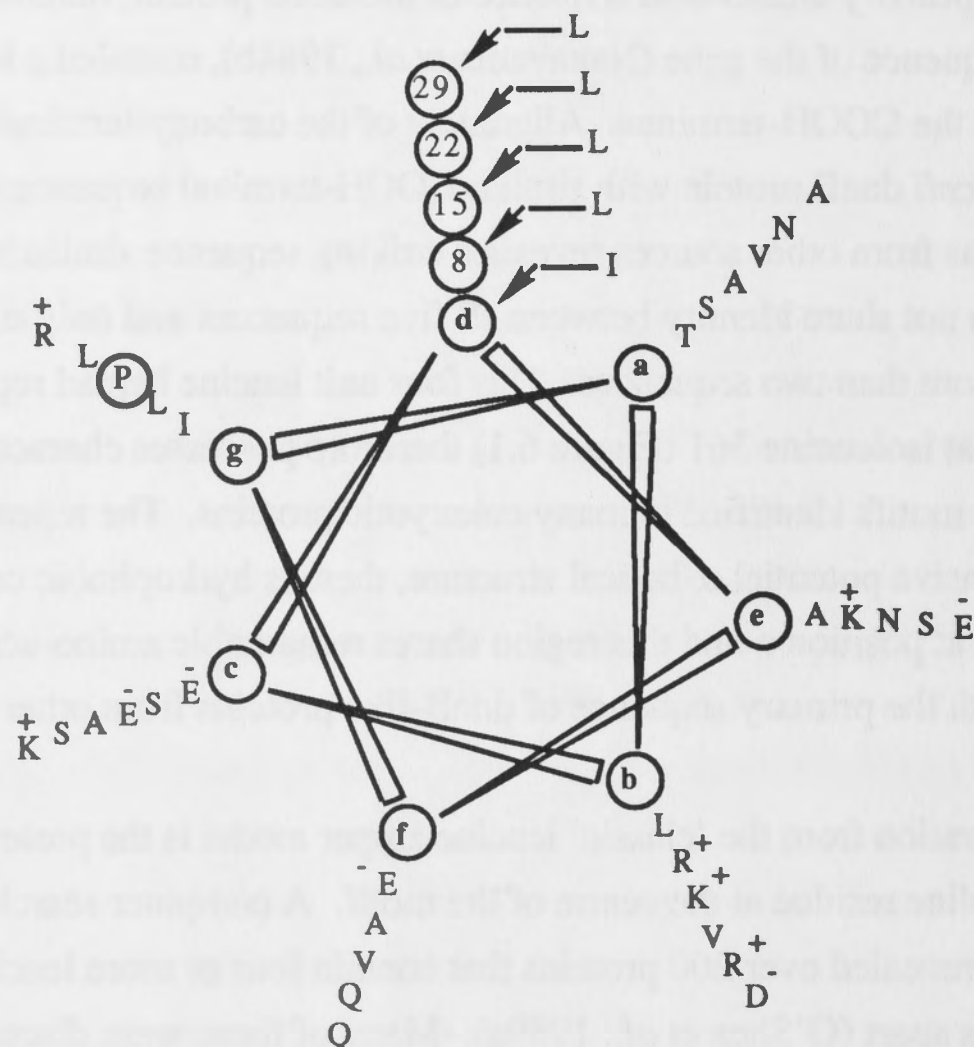


Figure 6.1

A. Helical wheel analysis of the leucine heptad repeat motif in dnaB. The view is end-to-end down the axis of a schematic α -helix. Each heptad repeat contributes two turns of helix, with individual positions in each heptad designated by the letters a through g. Leucine (circled) is usually found at position d (this repeat defines a leucine zipper motif). The amino terminal residue is Thr358 of *E. coli* dnaB and is placed at position a of the idealized helix. The circled proline is Pro378 and is central to the heptad repeat motif. Charged residues are also indicated.

B. A highly conserved amino acid sequence within dnaB and dnaB-related proteins from five separate sources. The leucine heptad repeat motif starts at isoleucine 361. Sources of protein sequences were; *E. coli* (Nakayama *et al.*, 1984b); *S. typhimurium* (Wong *et al.*, 1988); *C. trachomatis* (trachoma biovar; Sriprakash and Macavoy, 1987); *C. trachomatis* (LGV biovar; Hatt *et al.*, 1988); bacteriophage P22 (Backhaus and Petri, 1984). Amino-acid numbering is that of the *E. coli* sequence. Amino-acid abbreviations are the standard one-letter abbreviations (see Appendix B). Similarities are indicated by boxed amino acids: hatched = same amino acid; open = amino acids belonging to the same group (hydroxyl/small aliphatic: A, G, S, T; acid and acid amide: N, D, E, Q; basic: H, R, K; aliphatic: M, I, L, V; or aromatic: F, Y, W). The arrows indicate those leucines that form the leucine heptad repeat (in the d position), the asterisk indicates those amino acids at the +4 (a) position and the triangles a second hydrophobic repeat at the +3 (g) position.

B.

[illegible]

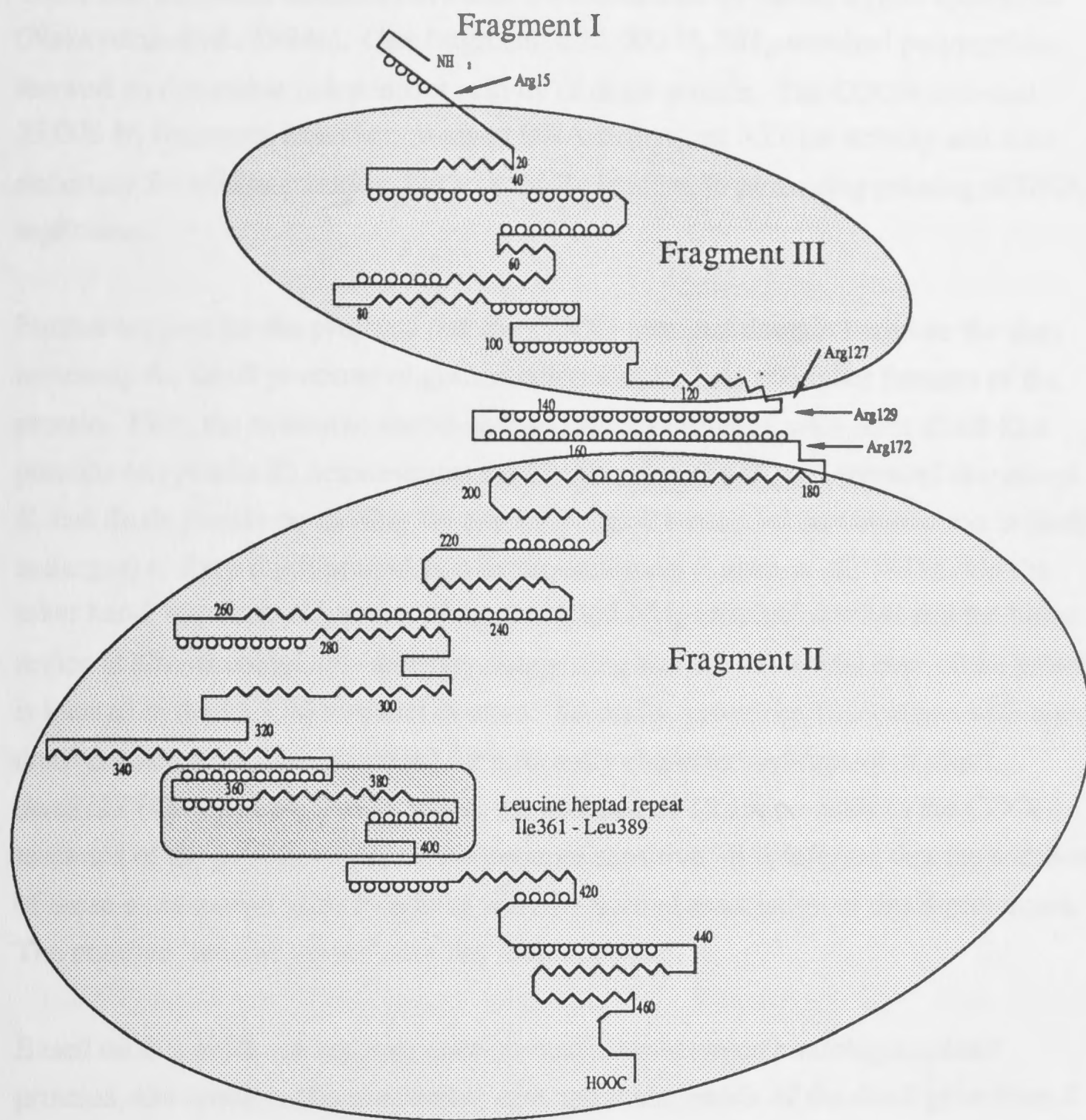


Figure 6.2

A model for secondary and domain structures of dnaB protein. Nakayama *et al.* (1984a) have proposed a scheme for domains of a dnaB protein monomeric unit derived from results of partial tryptic hydrolysis and secondary structure prediction from the primary amino acid sequence. The dnaB protein is proposed to comprise two discrete structural domains. The larger COOH-terminal domain ($M_r = 33$ kDa, Fragment II) maintains sites for DNA binding, nucleotide binding, DNA-dependent ATPase activity and oligomerization, whereas the smaller NH₂-terminal domain ($M_r = 12$ kDa, Fragment III) is required to maintain dnaC and primase interactions. Symbols used are: —, random coil; oooo, α helix; and \sim , β sheet. The arginine residues identified as the point of tryptic cleavage are shown. The more heavily shaded box denotes the leucine heptad repeat motif.

These two structural domains have also been identified by partial tryptic hydrolysis (Nakayama *et al.*, 1984a). One fragment, a 12,000 M_r NH_2 -terminal polypeptide, showed no detectable independent activity of dnaB protein. The $COOH$ -terminal 33,000 M_r fragment, however, retained DNA-dependent ATPase activity and sites necessary for oligomerization, but was totally inactive in promoting priming of DNA replication.

Further support for the proposal that the $COOH$ -terminal domain harbours the sites necessary for dnaB protomer oligomerization also include two other features of the protein. First, the extensive amino-acid sequence homology with other dnaB-like proteins (Appendix B) demonstrates the likelihood of the $COOH$ -terminal domain of *E. coli* dnaB protein containing the site for phage P1-encoded *ban* interaction (a dnaB analogue) to form dnaB protein-*ban* heteromultimers (Lanka *et al.*, 1978). On the other hand, sequence alignment of the predicted NH_2 -terminal domain and the hinge region is almost completely arbitrary, suggesting that the functional core of the protein is located in the $COOH$ -terminal domain. Secondly, protomer dissociation within the dnaB hexamer has been proposed for a recently characterized mutant of *dnaB*, *dnaB121* (Chang *et al.*, 1991), where the addition of 15 amino acids to the $COOH$ -terminus of the protein renders it temperature sensitive. It is inferred that the addition of these extra amino acids interferes with the normal association of dnaB protomers. The putative 'leucine zipper' also lies in this domain.

Based on this evidence and sequence comparisons between homologous dnaB proteins, site-specific oligonucleotide directed mutagenesis of the *dnaB* gene from *E. coli* was carried out with a view to determine the relationship between the leucine heptad repeat in the dnaB protein and protein oligomerization.

Site directed mutagenesis has become an important tool in probing the structural and functional significance of particular residues within a protein sequence. However, it is essential that amino-acid substitutions used to explore these aspects do not result in local or global perturbations in the protein that may result in conformational changes leading to incorrect conclusions on the nature of the substitution. Conserved or 'safe' substitutions are thus a requisite for the success of the mutant probe as an indicator of critical residues in structure and function. What was required in mutational analysis of the leucine heptad repeat in the dnaB protein was an amino-acid substitution that would maintain the α -helical and hydrophobic nature of the region but abrogate the putative protein-protein interface by removing the leucine residue requirement.

Valines, like isoleucine and methionine, are strong helix-forming residues (Chou and Fasman, 1978) and are therefore highly unlikely to block dnaB protein oligomerization through gross structural perturbation. Valine is also the most highly significant preferred exchange for leucine in actual tertiary structures of proteins (Bordo and Argos, 1991) and is the most commonly occurring residue found at the **a** position in characterized leucine zipper motifs. Leucine substitution by valine, therefore, should only affect the structural and functional characteristics of the dnaB protein if these highly conserved leucine residues are absolutely required. Valine was chosen as the substitution of choice.

Site directed mutagenesis studies of Jun, Fos and GCN4 (Gentz *et al.*, 1989; Turner and Tjian, 1989; Hu *et al.*, 1990) suggest that single leucine mutations may not affect dimerization of these proteins. Considering that the dnaB protein heptad repeat has the same unit length as that found in many of the eukaryotic transcription factors, it was desirable to construct not only single but also multiple valine-for-leucine substitutions in the same protein. This was made possible by mixing and hybridizing non-overlapping oligonucleotides with template DNA in a single mutagenesis reaction. For ease of detection, analysis and selection of single and multiple mutations, the formation and/or deletion of restriction endonuclease sites (without further effect on the protein sequence) were also directed by these oligonucleotides.

Mutagenesis was performed at three sites in the heptad repeat that were predicted to participate in oligomer formation. The mutant gene products were isolated and the oligomeric state of the mutant proteins and their ability to promote priming of DNA replication were assessed.

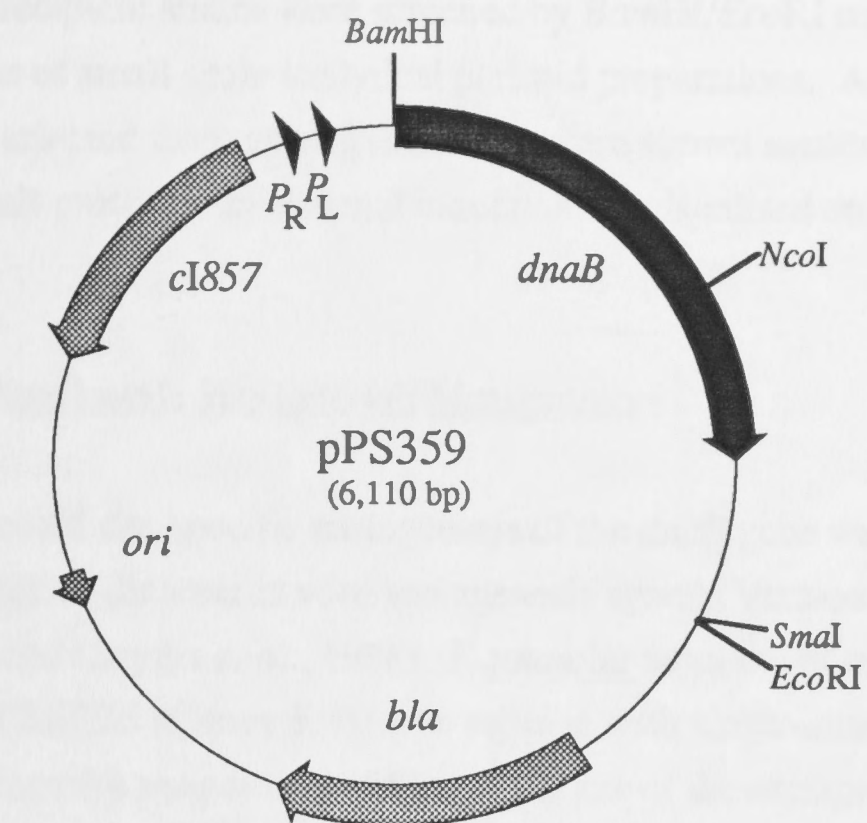
MATERIALS AND METHODS

6.2.1 Construction of the Phagemid pPS552

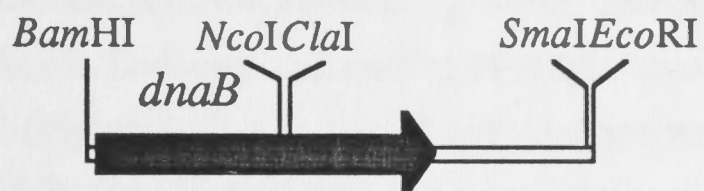
A phagemid derivative of pMA200U containing the intact *dnaB* gene was constructed by isolating the 2118-bp *Bam*HI/*Eco*RI *dnaB*-containing fragment from pPS359 (Section 4.2.3) and ligating this between the *Bam*HI and *Eco*RI sites in pMA200U (Figure 6.3). Competent AN1459 was transformed with the products of ligation and

Figure 6.3

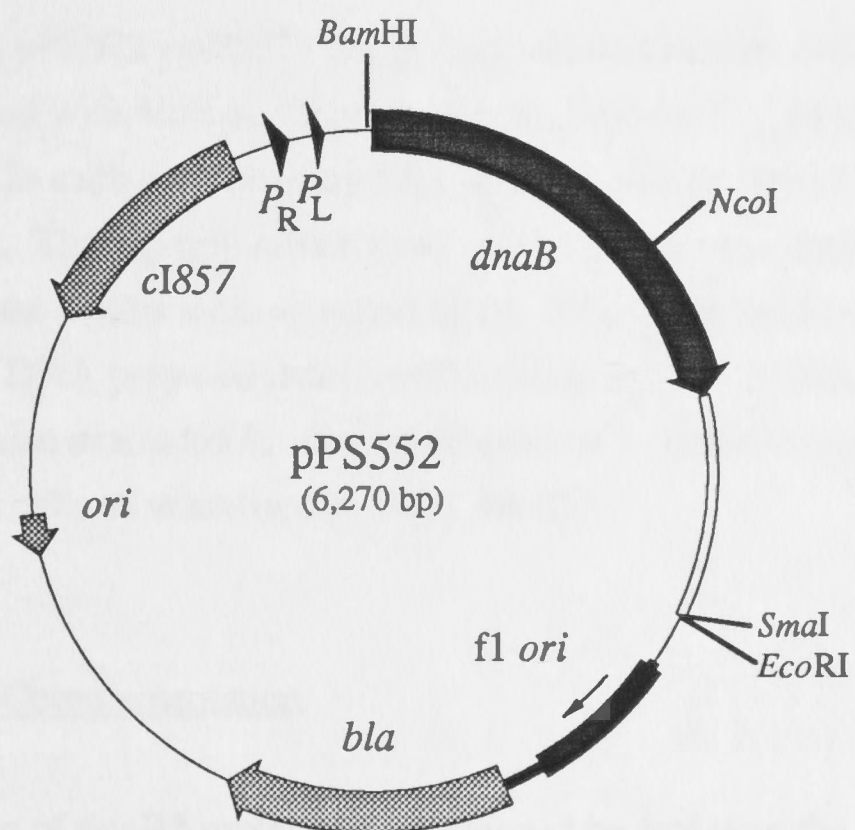
Construction of pPS552, a phagemid derivative of pMA200U bearing the *dnaB* gene (details of manipulations are in Section 6.2.1). Using this derivative, single-stranded DNA containing the coding strand of the *dnaB* gene was produced for use in oligonucleotide-directed site-specific mutagenesis (Section 6.2.2).



1. *Bam*HI/*Eco*RI
2. Fragment isolation



3. *Bam*HI/*Eco*RI-linearized pMA200U + ligase



ampicillin resistant recipient strains were screened by *Bam*HI/*Eco*RI restriction endonuclease digests of small-scale analytical plasmid preparations. A single plasmid (pPS552) was then selected from among those that were shown successfully to overproduce the *dnaB* protein after thermal induction, as visualised on SDS-PAGE.

6.2.2 Oligonucleotide Site-Directed Mutagenesis

Oligonucleotide-directed site-specific mutagenesis of the *dnaB* gene was performed using the oligonucleotide-directed *in vitro* mutagenesis system Version 2 (Amersham) essentially as described (Sayers *et al.*, 1988). Equimolar amounts of oligonucleotides RSC33, RSC34 and RSC35 (Figure 6.4) were utilized with single-stranded DNA template prepared from the phagemid pPS552. Products of the mutagenesis were used to transform competent AN1459 and strains were selected by ampicillin resistance at 30 °C. Successful mutagenesis was confirmed by restriction endonuclease digestion of plasmid DNA from small-scale analytical preparations with *Hae*II, *Dde*I and *Hpa*I. Successful overproduction of mutant *dnaB* gene products was verified by thermal induction of mutant bearing strains and identifying expressed gene products by SDS-PAGE. Dideoxy sequencing of single-stranded phagemid DNA using primer RSC32 (Figure 6.4) was carried out after preparing single-stranded DNA templates from AN2666 transformed with the selected phagemid mutants (Figure 6.8).

6.2.3 Construction of *dnaC/dnaB*(mutant) Co-overproduction Plasmids

The phagemid series pPS572 - pPS577 containing isolated mutant *dnaB* genes (*dnaB**) were digested with *Nco*I and *Eco*RI and the 1105-bp 3'-OH terminus of the *dnaB* gene isolated. In each case this fragment was ligated into *Nco*I/*Eco*RI linearized pPS562 (Figure 6.5). The ligation mixes were used to transform competent AN1459 and ampicillin resistant strains were screened by *Nco*I/*Eco*RI restriction endonuclease digestion of plasmid DNA prepared from small-scale analytical preparations. Selected transformants were also evaluated for overproduction of both *dnaC* and *dnaB* proteins from thermoinduced cells as visualised by SDS-PAGE.

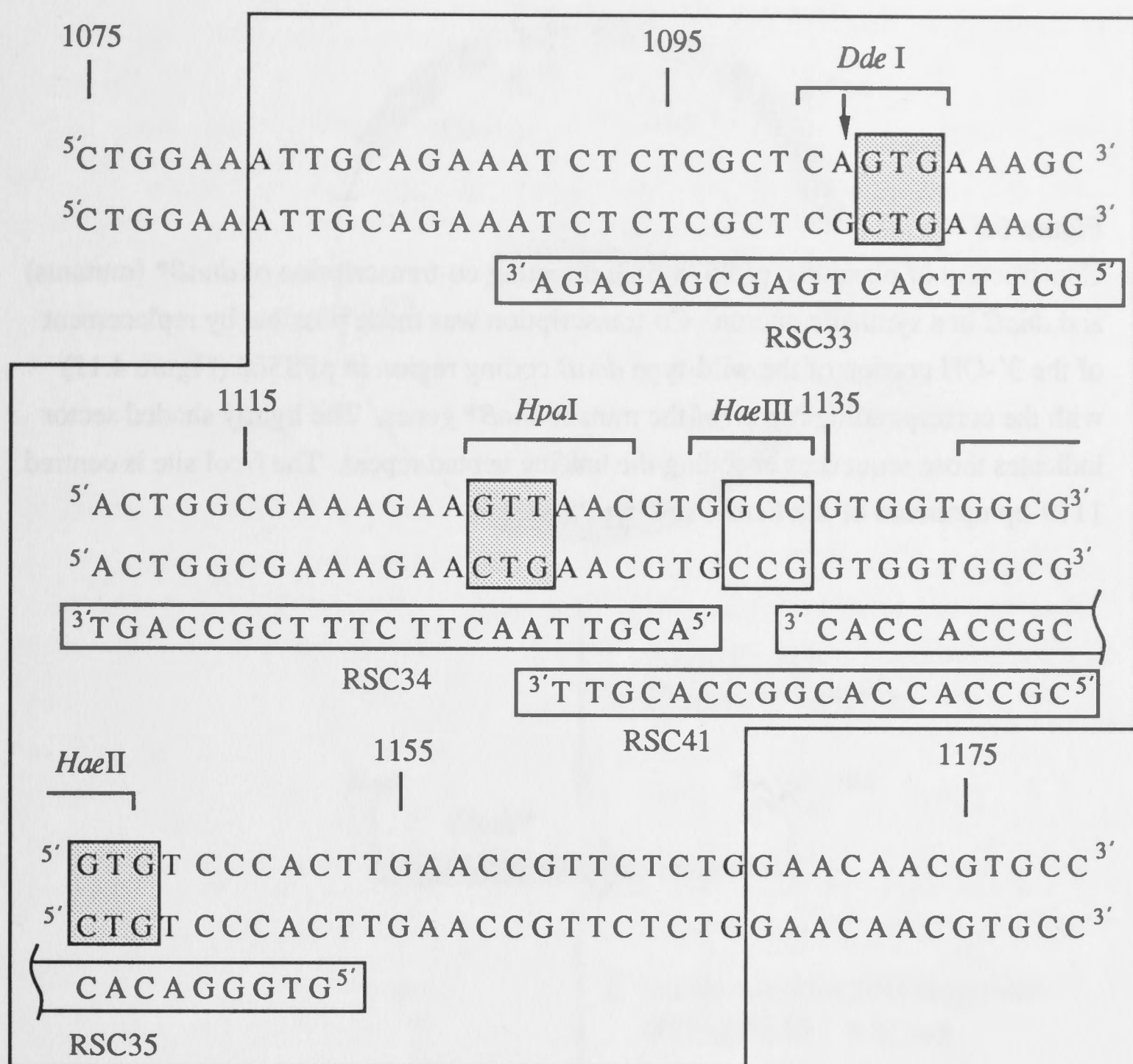
6.2.4 *dnaB* Complementation

Low-level production of *dnaB** proteins was achieved by isolating the 2124-bp *dnaB** containing *Bam*HI/*Eco*RI fragments from the phagemid series pPS572 - pPS577 and

Figure 6.4

A strategy for site-directed mutagenesis of the leucine heptad repeat in *E. coli dnaB*. Oligonucleotides RSC 33, 34 and 35 (boxed 3' → 5' sequences) were synthesized for site-directed mutagenesis of nucleotides encoding leucine residues at positions 368, 375 and 382 of the *E. coli dnaB* protein (shaded boxes). Shown are sequences of the *dnaB* gene from nucleotide 1075 to 1249 (numbered with respect to the 5' ATG start codon of the gene). The lower sequence is that of the wild-type *dnaB* and the upper sequence is the expected outcome from successful mutagenesis. The position of hybridization of the oligonucleotide RSC32 used to prime dideoxy sequencing reactions is shown. RSC41 is a synthetic oligonucleotide designed to alter the sequence coding for a unique proline residue at position 378 of the *E. coli dnaB* protein to that of an alanine codon. The oligonucleotides directing mutagenesis also direct the formation or deletion of a restriction endonuclease site in the *dnaB* gene. Following mutagenesis, the mutant products were screened by analysis of restriction endonuclease cleavage patterns. The mutation and the resulting appearance or concurrent loss of restriction endonuclease site are given in the table below.

Oligonucleotide	Mutation	Restriction endonuclease site affected (+ or -)	Loss of Fragment	Fragments Gained
RSC33 (19-mer)	L368V	+ <i>DdeI</i>		1239 2019 → + 780
RSC34 (21-mer)	L375V	+ Unique <i>HpaI</i>		Linearizes plasmid DNA
RSC35 (18-mer)	L382V	- <i>HaeII</i>	1280 + → 1367 87	
RSC41 (18-mer)	P378A	+ <i>HaeIII</i>		199 381 → + 182



*Nci*II 1195

5' GACAAACGCCCCGGTCAACTCCGACCTGCGTGAATC 3'

5' GACAAACGCCCCGGTCAACTCCGACCTGCGTGAATC 3'

1215 | 1235

5' TGGCTCTATCGAGCAGGATGCGGACTTGATCATGT 3'

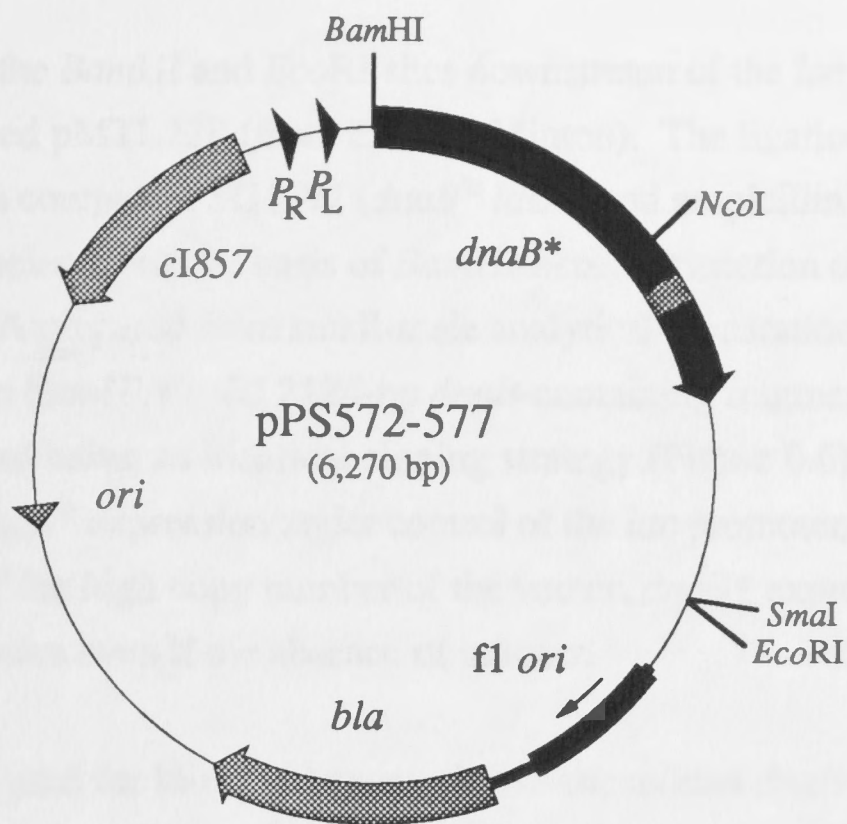
5' TGGCTCTATCGAGCAGGATGCGGACTTGATCATGT 3'

3' GATAGCTCGTCCTACGC 5'

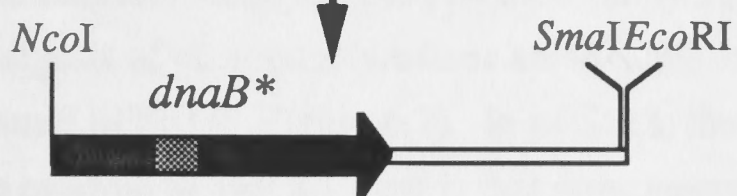
RSC32

Figure 6.5

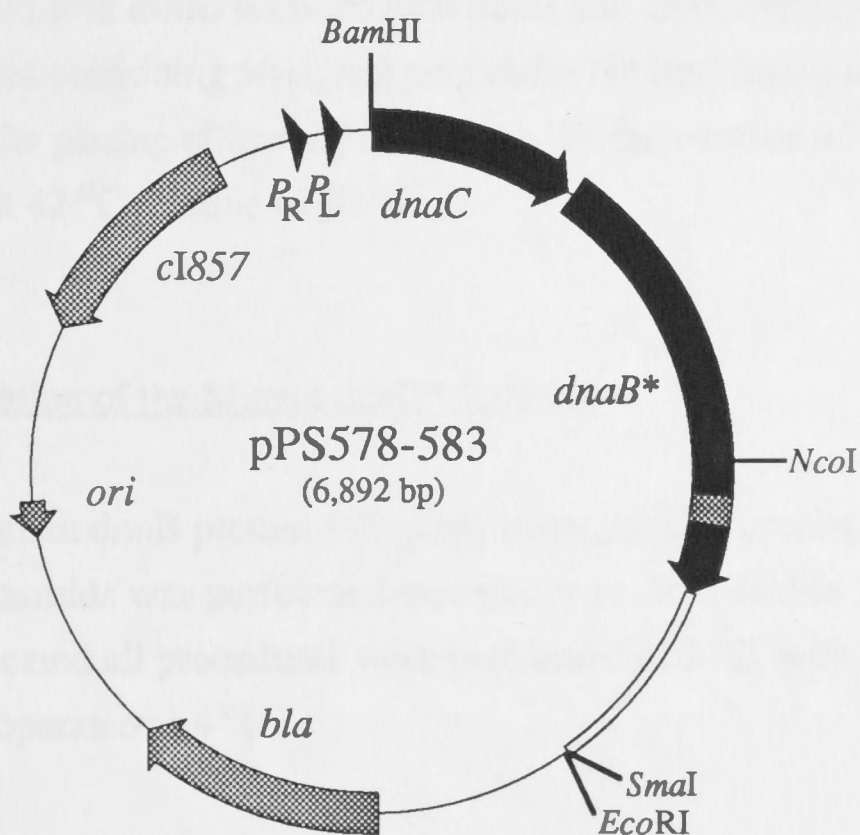
Construction of plasmids, pPS578-583, directing co-transcription of *dnaB** (mutants) and *dnaC* in a synthetic operon. Co-transcription was made possible by replacement of the 3'-OH portion of the wild-type *dnaB* coding region in pPS562 (Figure 4.11) with the corresponding region of the mutant *dnaB** genes. The lightly shaded sector indicates those sequences encoding the leucine heptad repeat. The *Nco*I site is centred 1119-bp upstream of the *Eco*RI site in pPS562.



1. *Nco*I/*Eco*RI
2. Fragment isolation



3. Large *Nco*I/*Eco*RI fragment from pPS562 + ligase



ligating each between the *Bam*HI and *Eco*RI sites downstream of the *lac* promoter in *Bam*HI/*Eco*RI linearized pMTL22P (from Dr N.P. Minton). The ligation reactions were used to transform competent SG1692 (*dnaB*^{ts} *lacI*⁺) and ampicillin resistant recipient strains were selected on the basis of *Bam*HI/*Eco*RI restriction endonuclease digests of plasmid DNA prepared from small-scale analytical preparations. A control plasmid harbouring the *Bam*HI/*Eco*RI 2124-bp *dnaB*-containing fragment from pPS359 was constructed using an identical cloning strategy (Figure 6.6). These manipulations place *dnaB** expression under control of the *lac* promoter/operator of the vector. Because of the high copy number of the vector, *dnaB** expression is expected to be constitutive even if the absence of inducer.

A second plasmid designed for low-level expression of the mutant *dnaB** protein L368V·L375V was constructed by isolating the 2431-bp *Kpn*I/*Eco*RI fragment containing the intact mutant *dnaB** gene from pPS580 (Figure 6.5). This was ligated between the *Kpn*I and *Eco*RI sites downstream of the *lac* promoter in pMTL23P (from Dr N.P. Minton). The ligation reaction mixture was used to transform competent SG1692 and an ampicillin resistant strain selected on the basis of *Kpn*I/*Eco*RI restriction endonuclease digests of plasmid DNA from small-scale analytical preparations (giving plasmid pPS618; Figure 6.7). In pPS618, the only portion of *dnaB* remaining from the original mutant plasmid is that downstream of the *Nco*I site. A control plasmid harbouring the *Kpn*I/*Eco*RI 2431-bp *dnaB* containing fragment from pPS562 was constructed using an identical cloning strategy (Figure 6.7).

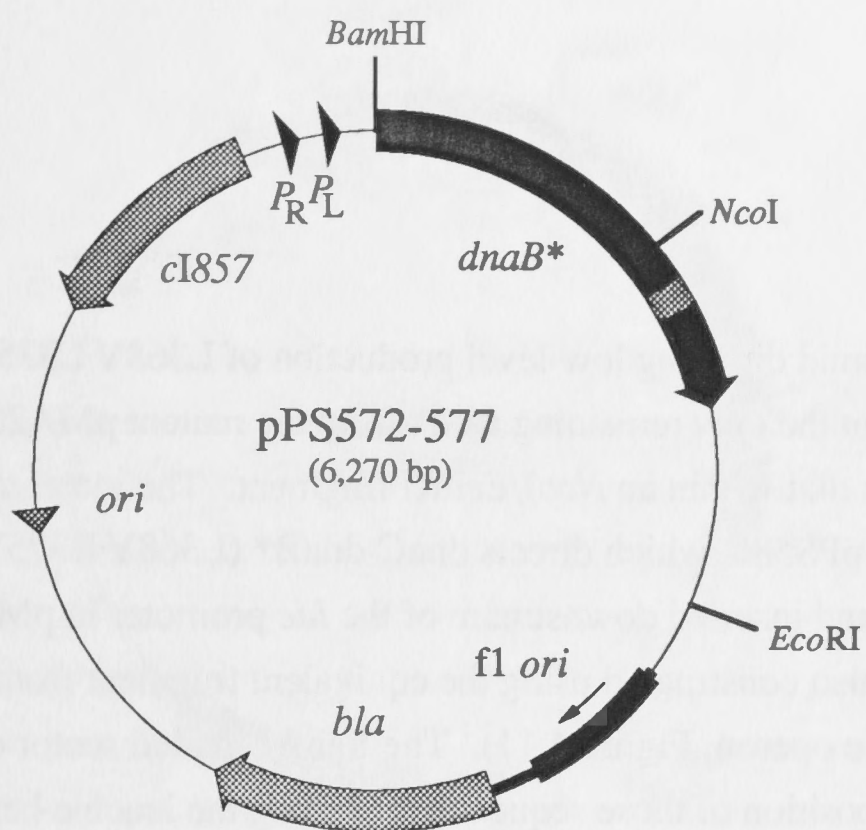
Complementation efficiency was determined by growing selected strains in LBT containing 50 µg/ml ampicillin to an $A_{595} = 0.5$, then chilling the cells suspensions quickly. A series of ten-fold dilutions were then made and cells were spread in duplicate on LBT plates containing 50 µg/ml ampicillin. Plates were then incubated at 30 °C and 42 °C and the plating efficiency determined as the number of colonies appearing after 16 h at 42 °C relative to 30 °C.

6.2.5 Purification of the Mutant *dnaB** Proteins

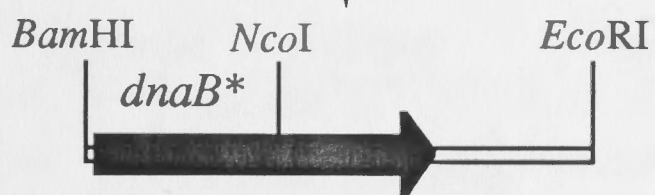
The purification of mutant *dnaB* protein following expression of mutant *dnaB** genes in pPS562-derived plasmids was performed essentially as described in Section 5.2.2. Unless otherwise indicated all procedures were performed at 0 °C, with dialysis and column purifications operated at 4 °C

Figure 6.6

Low-level expression of the mutant *dnaB** genes, engineered by site-directed mutagenesis, was achieved by isolating the intact *dnaB** genes fused to the synthetic ribosome-binding site of the vector within *Bam*HI/*Eco*RI cassettes and inserting these downstream of the *lac* promoter in pMTL22P. This placed *dnaB** expression under the control of the *lac* promoter/operator of the vector. The identical fragment of pPS359 (Figure 4.7) was similarly treated to engineer a control plasmid. The lightly shaded sector of the *dnaB** gene indicates the position of those sequences encoding the leucine heptad repeat motif.



1. *Bam*HI/*Eco*RI
2. Fragment isolation



3. *Bam*HI/*Eco*RI-linearized pMTL22P + ligase

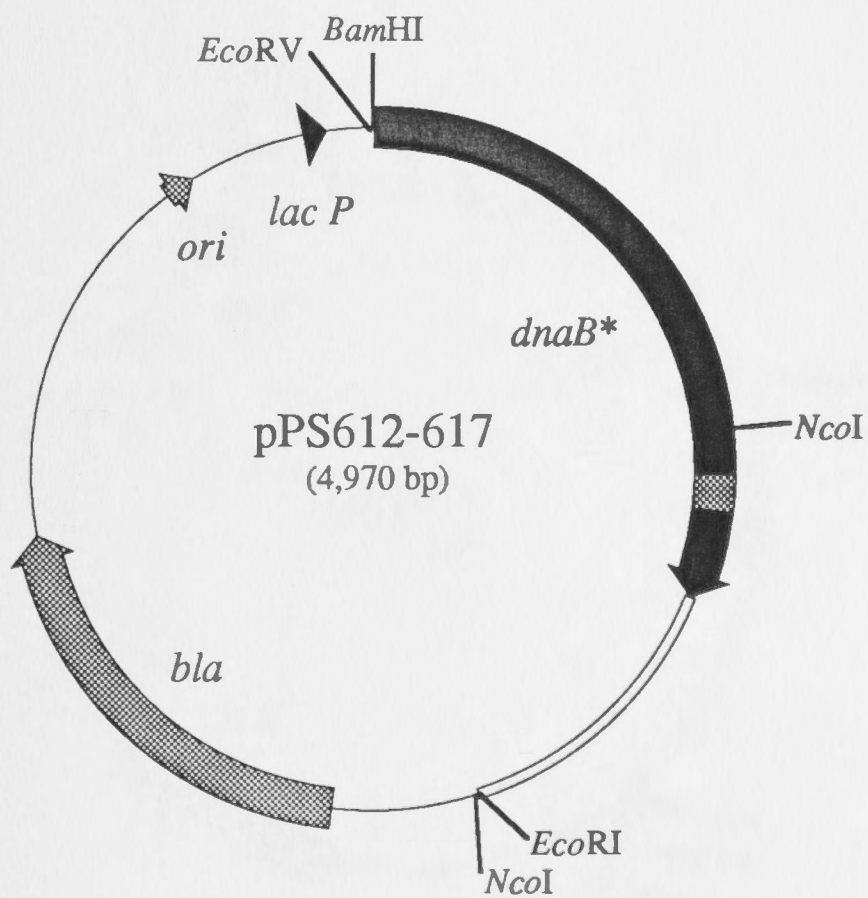
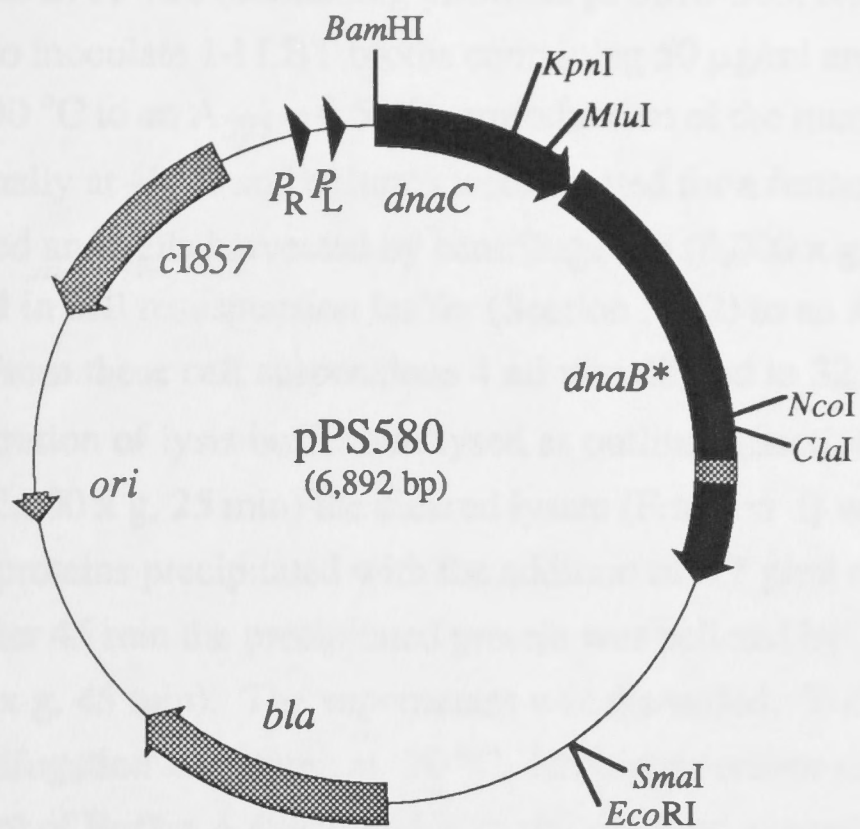
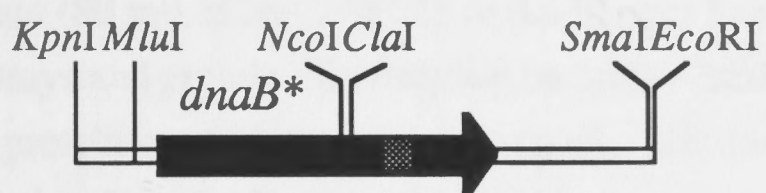


Figure 6.7

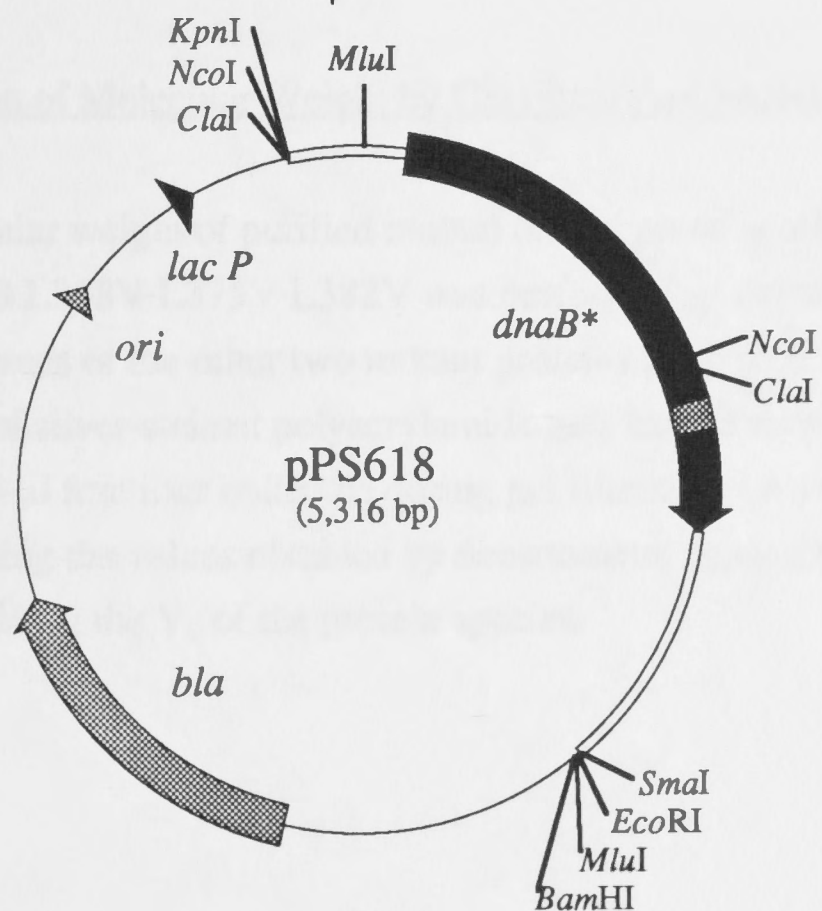
A second plasmid directing low-level production of L368V·L375V was constructed in such a way that the only remaining DNA from the mutant pMA200U derivative (pPS574) was that within an *NcoI*/*EcoRI* fragment. The intact *dnaB** gene was isolated from pPS580, which directs *dnaC*-*dnaB** (L368V·L375V) overproduction (Figure 6.5), and inserted downstream of the *lac* promoter in pMTL23P. A control plasmid was also constructed using the equivalent fragment from pPS562 (the *dnaB*-*dnaC* synthetic operon, Figure 4.11). The lightly shaded sector of the *dnaB** gene indicates the position of those sequences encoding the leucine heptad repeat motif.



1. *Kpn*I/*Eco*RI
2. Fragment isolation



3. *Kpn*I/*Eco*RI-linearized pMTL23 + ligase



Overnight cultures of RSC711-716 (containing plasmids pPS578-583, respectively; Figure 6.5) were used to inoculate 1-l LBT broths containing 50 $\mu\text{g/ml}$ ampicillin. These were aerated at 30 °C to an $A_{595} = 0.5$. Overproduction of the mutant proteins was then induced thermally at 42 °C and cultures were aerated for a further 4 h. Cell suspensions were chilled and cells harvested by centrifugation (8,000 x g, 10 min). Cells were resuspended in cell resuspension buffer (Section 5.2.2) to an $A_{595} = 200$ and stored at -70 °C. From these cell suspensions 4 ml was diluted to 32 ml ($A_{595} = 25$) to the final concentration of lysis buffer and lysed as outlined (Section 5.2.2). After centrifugation (12,000 x g, 25 min) the cleared lysate (Fraction I) was decanted and the soluble dnaB* proteins precipitated with the addition of 0.2 g/ml of ammonium sulfate. After 45 min the precipitated protein was pelleted by centrifugation (40,000 x g, 45 min). The supernatant was discarded. The pellets were repacked by centrifugation and stored at -70 °C. Each ammonium sulfate pellet was resuspended in 5 ml of Buffer A (Section 5.2.2) and dialysed against 2 x 1-l of this buffer over 6 h. The dialysed samples were then diluted to 10 ml with Buffer A (Fraction II) and loaded onto a DEAE-cellulose column (1 x 10 cm) pre-equilibrated with Buffer A. The column was washed with 16 ml of Buffer A and the dnaB* protein eluted in a gradient (80 ml) of NaCl (0.025 to 0.6 M) at a flow rate of 6 ml/h. Bradford dye-binding assays and protein visualization by SDS-PAGE were used to determine which dnaB* protein-containing fractions to pool. The dnaB* proteins pooled following purification from RSC strains 713 and 715 were concentrated by a step-wise elution from an identical DEAE-cellulose column with 0.4 M NaCl in Buffer A. Pooled fractions from all six purifications were stored at -70 °C.

6.2.6 Estimation of Molecular Weight by Gel Filtration Chromatography

Estimation of the molecular weight of purified mutant dnaB* proteins other than dnaB L368V·L375V and dnaB L368V·L375V·L382V was performed as described (Section 2.4.5). The elution volumes of the other two mutant proteins were determined by densitometric scanning of silver-stained polyacrylamide gels loaded with equivalent volumes taken from 250- μl fractions collected during gel filtration. An elution profile was constructed by plotting the values obtained by densitometry against the elution volume of fractions, yielding the V_e of the protein species.

RESULTS

Site-directed mutagenesis

Site-specific mutants in the COOH-terminal leucine heptad repeat were made utilizing single-stranded DNA prepared from a *dnaB*-containing derivative of pMA200U (Figure 2.1). This phagemid vector contains the viral strand origin of the filamentous bacteriophage f1 (Elvin *et al.*, 1989). It was therefore possible to prepare single-stranded DNA in the M13-sensitive (F^+) host strain AN2666 with provision of the M13 gene II product *in trans* by infection with the M13 derivative M13K07 (Mead *et al.*, 1986).

Mutagenesis yielded a collection of mutants which included single, double and triple sequence alterations (Figure 6.8), demonstrating that it is possible to utilize multiple oligonucleotides in a single reaction. The use of linked restriction endonuclease sites as markers for mutagenesis proved invaluable in the quick selection of desired mutants. Of the 113 transformants screened, 80 % had at least one mutation and 25 % had multiple site-directed alterations. Unfortunately, of the seven possible mutational combinations predicted from this experiment, only six were retrieved. These were given the plasmid and strain designations described (Table 6.1). The mutant gene products from all six strains were overproduced successfully following thermal induction (Figure 6.9).

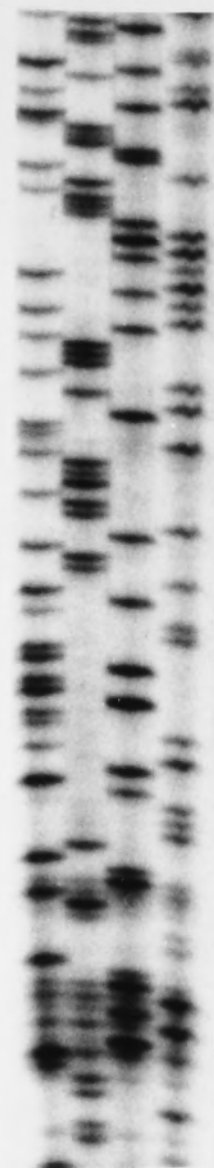
Dideoxy sequencing of these mutants confirmed that the expected alterations had occurred and that the genes were otherwise intact (Figure 6.8). The exception, the L382V mutant from pPS575, revealed a second mutation at nucleotide 1101 in *dnaB*. This mutation was an inconsequential alteration of the Ser382 codon (TCG to TCA), presumably a partial site-specific mutation directed by oligonucleotide RSC33. This was not detected during restriction endonuclease screening of plasmid DNA as the complete alteration of the sequence to provide a *Hae*II site marker had not occurred.

In light of earlier experiments addressing solubility of overproduced *dnaC* and *dnaB* proteins (Chapter 4 and 5) it was considered necessary to express the mutant proteins in the presence of *dnaC*. This was achieved simply by replacing the 3'-OH terminus of the *dnaB* gene in the *dnaC/dnaB* synthetic operon, pPS562, with the equivalent sequence containing all mutations from the altered *dnaB** genes. Proteins overproduced from these sources (Figure 6.9) were used as a source for the purification of the mutant proteins.

Figure 6.8

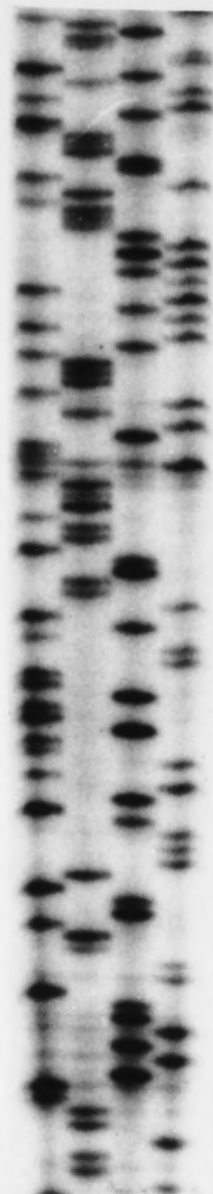
Dideoxy nucleotide sequencing of derivatives of pPS552 (pMA200U *dnaB*⁺). The preparation of template DNA (Section 2.3.12) and the sequencing reactions (Section 2.3.14) using primer RSC32 (Figure 6.4) were as described. The figure shows the complete wild-type DNA sequence in the region encoding the leucine heptad repeat of *dnaB* (pPS552). Other sequences are labelled only where substitutions have occurred with respect to the primary sequence.

pPS552
G A T C



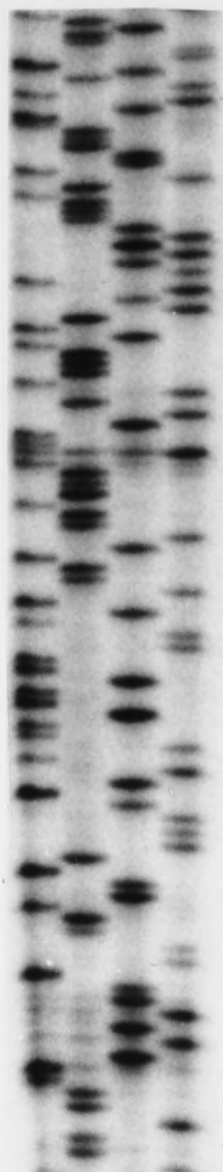
GA
AA
CC
GT
AC
CG
TG
GA
TT
GC
AA
AT
CT
CG
CT
CG
CT
GA
AA
GC
AC
TG
GC
AA
GA
AC
TG
AA
CG
TG
CC
GG
TG
GG
CG
TT
CC
CA
GT
TG
AA
CC
GT
TC
TG
GA
CA
CA

pPS572
G A T C



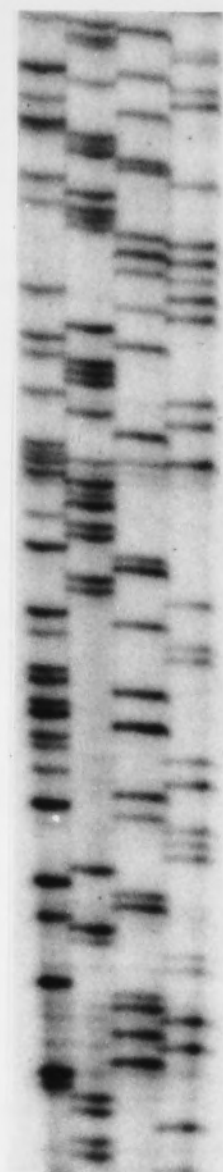
G
A
G
T
T
A
A
C

pPS573
G A T C



T
C
A
G
T
G
A
A
A

pPS574
G A T C



T
C
A
G
T
G
A
A
A
G
A
G
T
T
A
A
C

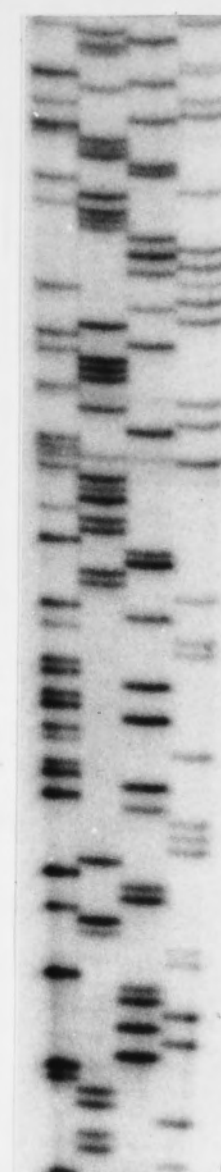
pPS575
G A T C



T
C
A
C
T
G
A
A
A

G
G
G
G
G
G
G
G
G
G
G
G
G

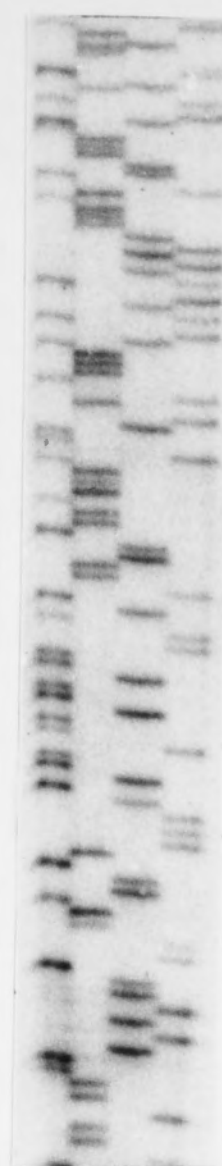
pPS576
G A T C



T
C
A
G
T
G
A
A
A
G
A
G
T
T
A
A
C

G
G
G
G
G
G
G
G
G
G
G
G
G

pPS577
G A T C



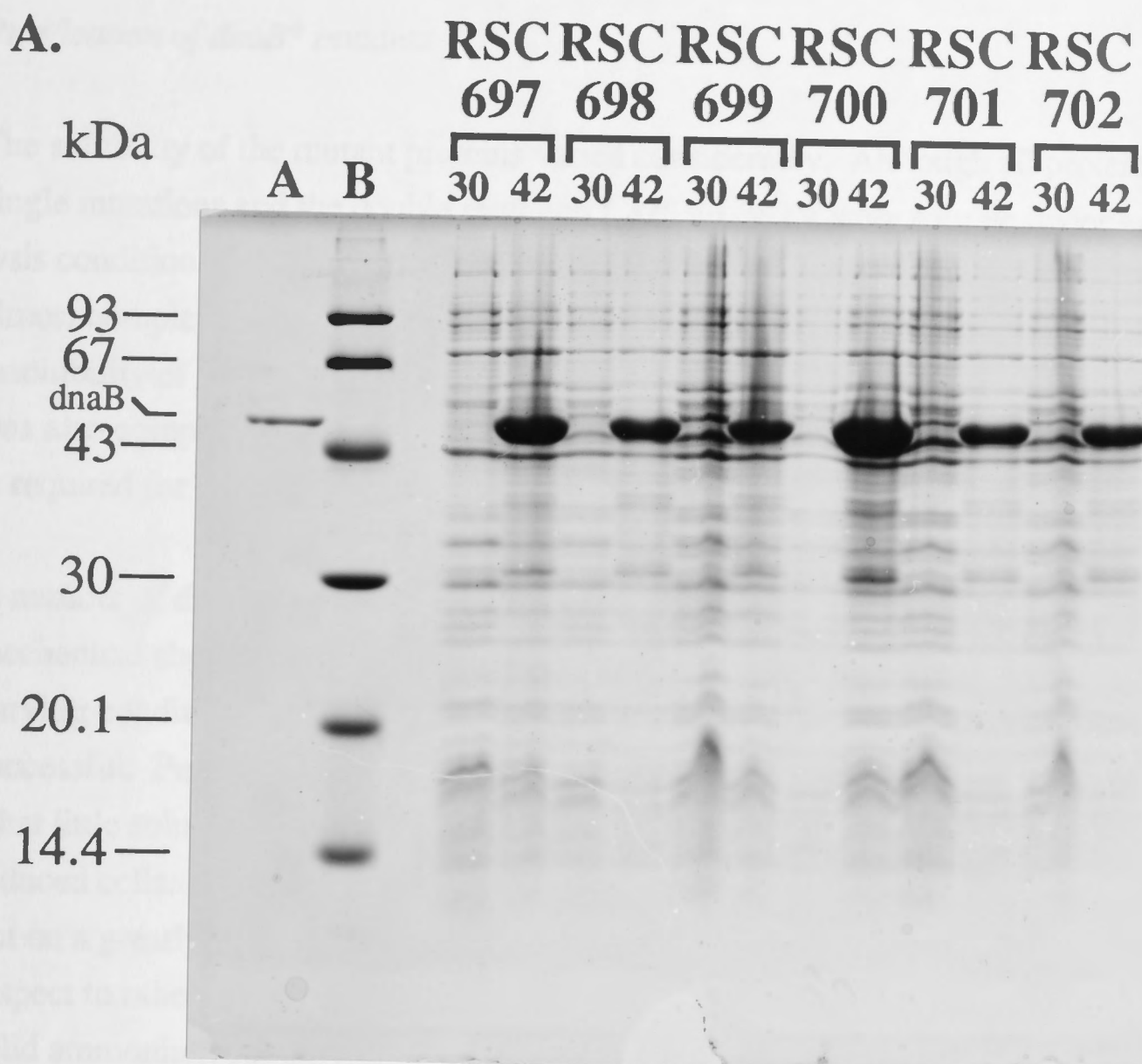
G
A
G
T
T
A
A
C

G
G
G
G
G
G
G
G
G
G
G
G
G

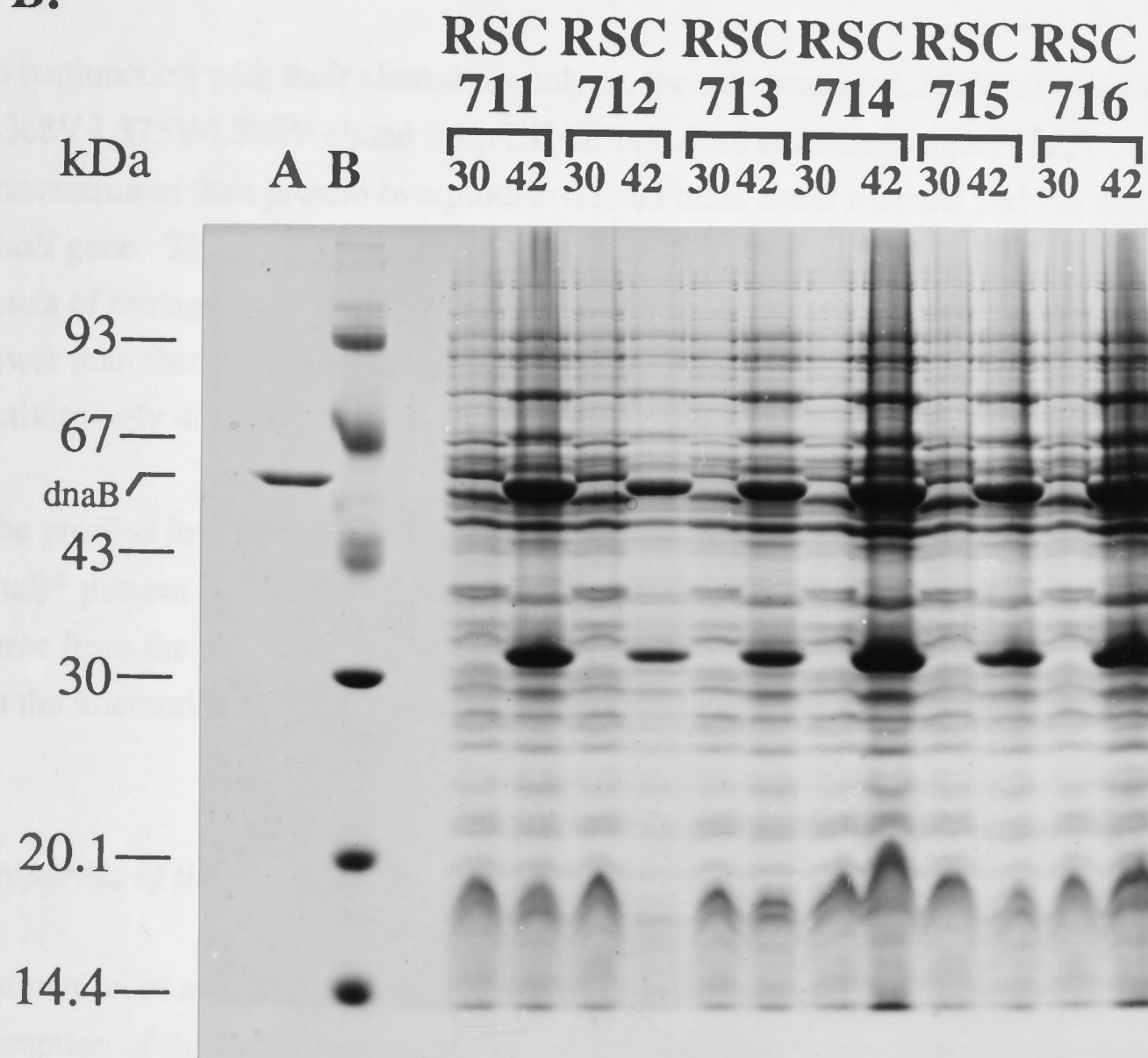
Figure 6.9

Overproduction of dnaB* proteins directed by plasmids engineered by mutagenesis of pPS552, visualized by SDS polyacrylamide gel electrophoresis. Cultures (20 ml) of strains RSC697-702 (A) and RSC711-716 (B) were grown at 30 °C to $A_{595} \sim 0.5$. Cells were harvested from a portion (1 ml), and the remainder of the culture was shaken at 42 °C. Further 1-ml samples were removed after 4 h. Cells from the 1-ml portions of cultures removed prior to (30) and after 4 h of treatment at 42 °C (42) were resuspended to $A_{595} = 10$ in an SDS-gel loading buffer and heated for 2 min at 95 °C. Portions (20 μ l) were loaded onto two 12.5-% polyacrylamide gels. Protein markers (sizes in kDa; lane B) were as in Section 2.4.3 and an authentic sample of dnaB protein ($\sim 4 \mu$ g; lane A) is indicated.

A.



B.



Purification of dnaB mutants*

The solubility of the mutant proteins varied considerably. Although all proteins with single mutations and the double mutation L375V·L382V were soluble under gentle lysis conditions, both the triple mutant and the double mutant L368V·L375V were almost completely insoluble. One notable observation made in conjunction with the insolubility of these two mutants was that in both cases the overproduced dnaC protein was also completely insoluble. This supports the proposal that the presence of dnaB is required for the solubilization of the *dnaC* gene product and *vice versa* (Chapter 5).

A number of different lysis conditions were trialed to solubilize the proteins, including mechanical shearing, use of detergents, freeze thawing and lysozyme lysis under varying conditions of pH, ionic strength and temperature. However, none was successful. Purification of the insoluble mutant proteins was therefore performed on what little soluble material was present following a gentle lysozyme lysis of thermally induced cells. The purification procedure was essentially as described (Section 5.2.2) but on a greatly reduced scale. Each mutant protein was purified to >95 % with respect to other proteins (Figure 6.11) in a two step process: (i) precipitation with solid ammonium sulfate from a soluble cell extract; and (ii) DEAE-cellulose anion-exchange chromatography (Figure 6.10).

In conjunction with their altered phenotype, the two mutants L368V·L375V and L368V·L375V·L382V eluted from DEAE-cellulose at a much higher NaCl concentration than protein overproduced from other *dnaB* mutants and the wild-type *dnaB* gene. This probably reflects an altered conformation in the two proteins as a result of mutagenesis. Specific activities from these two proteins were considerably lower than that of the other mutants with L368V·L375V·L382V having a specific activity only 40.5 % of the wild-type protein (Table 6.1).

The profiles for DEAE-cellulose also suggest some variation in the properties of dnaB* present in dnaB*-dnaC complexes (*cf.* Chapter 5). The protein analysed here came from the peaks of free dnaB*. More work is required before further comment on the interaction of the mutant dnaB* protein with dnaC is warranted.

Properties of the dnaB mutants

Estimation of molecular weight by gel filtration indicated that all mutants, with the exception of the triple mutant, were similar in this respect to the non-mutant protein, a

Figure 6.10

Separation on DEAE-cellulose of dnaB* proteins from thermally-induced strains (RSC711 - RSC716) containing plasmids which direct expression of site-directed mutant proteins. Conditions for purification are as described in Section 6.2.5. The resuspended and dialysed ammonium sulfate pellets of the soluble dnaB* proteins remaining in the cleared cell lysate were diluted to 10 ml with Buffer A and applied to columns (1 x 10 cm) of DEAE-cellulose anion-exchange resin (DE-52) pre-equilibrated in Buffer A. The column was washed with Buffer A (16 ml) and bound proteins were eluted with a linear gradient (80 ml) of 0.025 to 0.6 M NaCl in Buffer A at a flow rate of 6 ml/h. Fractions (1 ml each) were collected and Bradford dye-binding assays and protein visualization by SDS-PAGE were used to determine which dnaB* protein containing fractions to pool. The lower panel (A) shows the elution profile of the four soluble dnaB* proteins and the top panel (B) is that of the "insoluble mutants". Peak fractions identified as containing dnaB* protein (not shown) are marked with arrows.

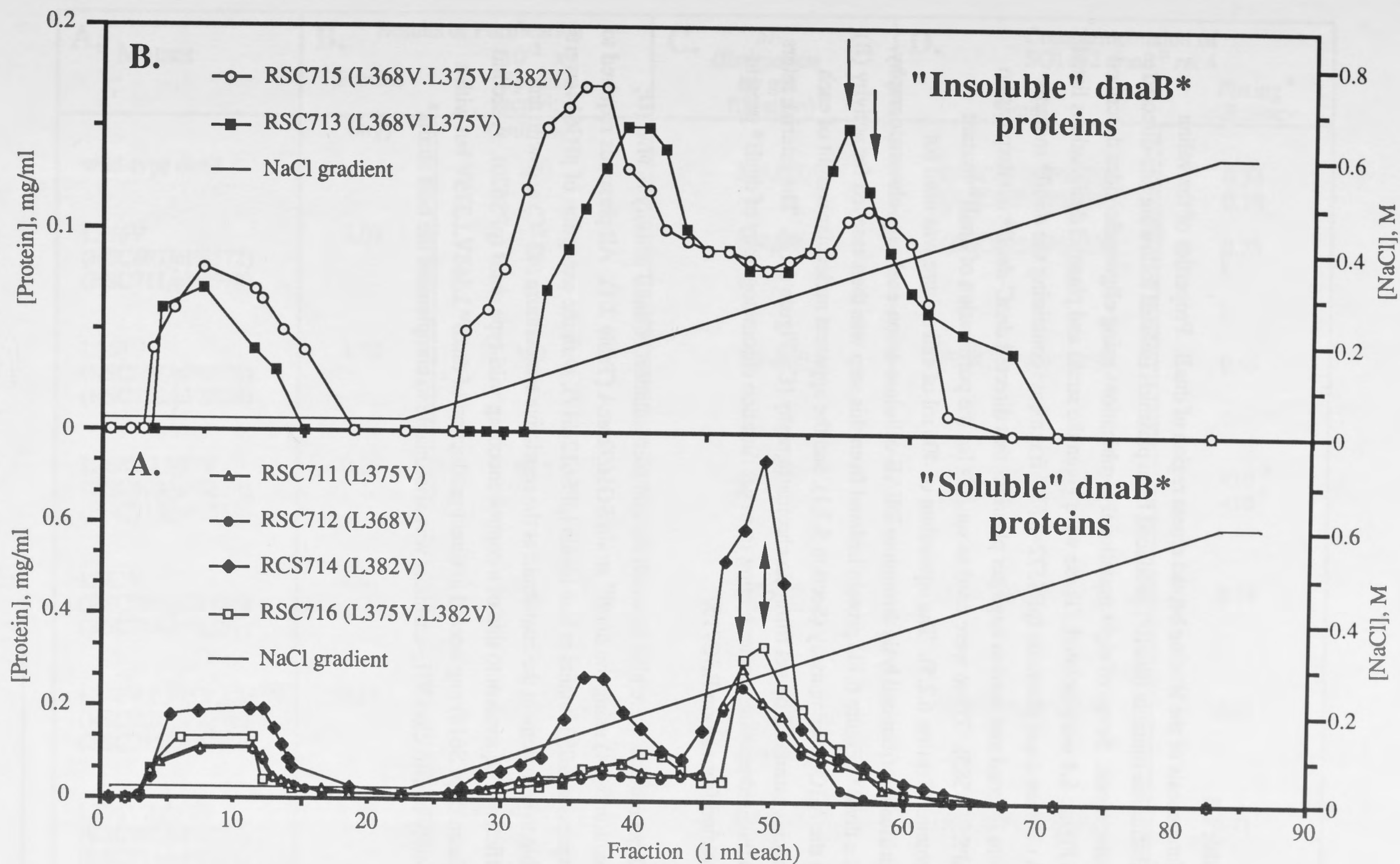


Table 6.1

Mutagenesis of the leucine heptad repeat region of *dnaB*. Properties of the valine substitution mutants (*dnaB**) produced from plasmids isolated following site-directed mutagenesis. Seven of eight mutational combinations using oligonucleotides described in Figure 6.4 were isolated. These were given the strain and plasmid designations listed (A). From these plasmids (pPS572–577), fragments containing the *dnaB** mutants were isolated and used to engineer plasmids that directed *dnaC-dnaB** overexpression (pPS578–583). These were used as sources for the purification of *dnaB** mutant proteins (Section 6.2.5). The equivalent of 400 ml of cell culture was used for purification from cell lysis through to DEAE-cellulose anion-exchange chromatography. The *dnaB** (Figure 6.11) protein isolated from this step was then assayed for activity (B) in the ABC-priming assay (Section 5.2.1), and the apparent molecular weight of each species estimated by gel filtration chromatography (C; Figure 6.12). The asterisk refers to the predominant species eluting from gel filtration chromatography of *dnaB** protein isolated from strain RSC715.

These mutants were also assessed for complementation of *dnaB* activity *in vivo* (D; Section 6.2.4) using the *dnaB*^{ts} strain SG1692*recA* (Table 2.1). All plasmids required to express *dnaB** mutants at low levels (pPS612–617), with the exception of pPS614, were able to complement the host strain at the restrictive temperature (42 °C) and with an efficiency equivalent to that of a control encoding wild-type *dnaB* (pPS620). A second plasmid (pPS618) engineered to direct production of *dnaB** L368V.L375V but with a newly constructed NH₂-terminus was also unable to complement the host strain.^a

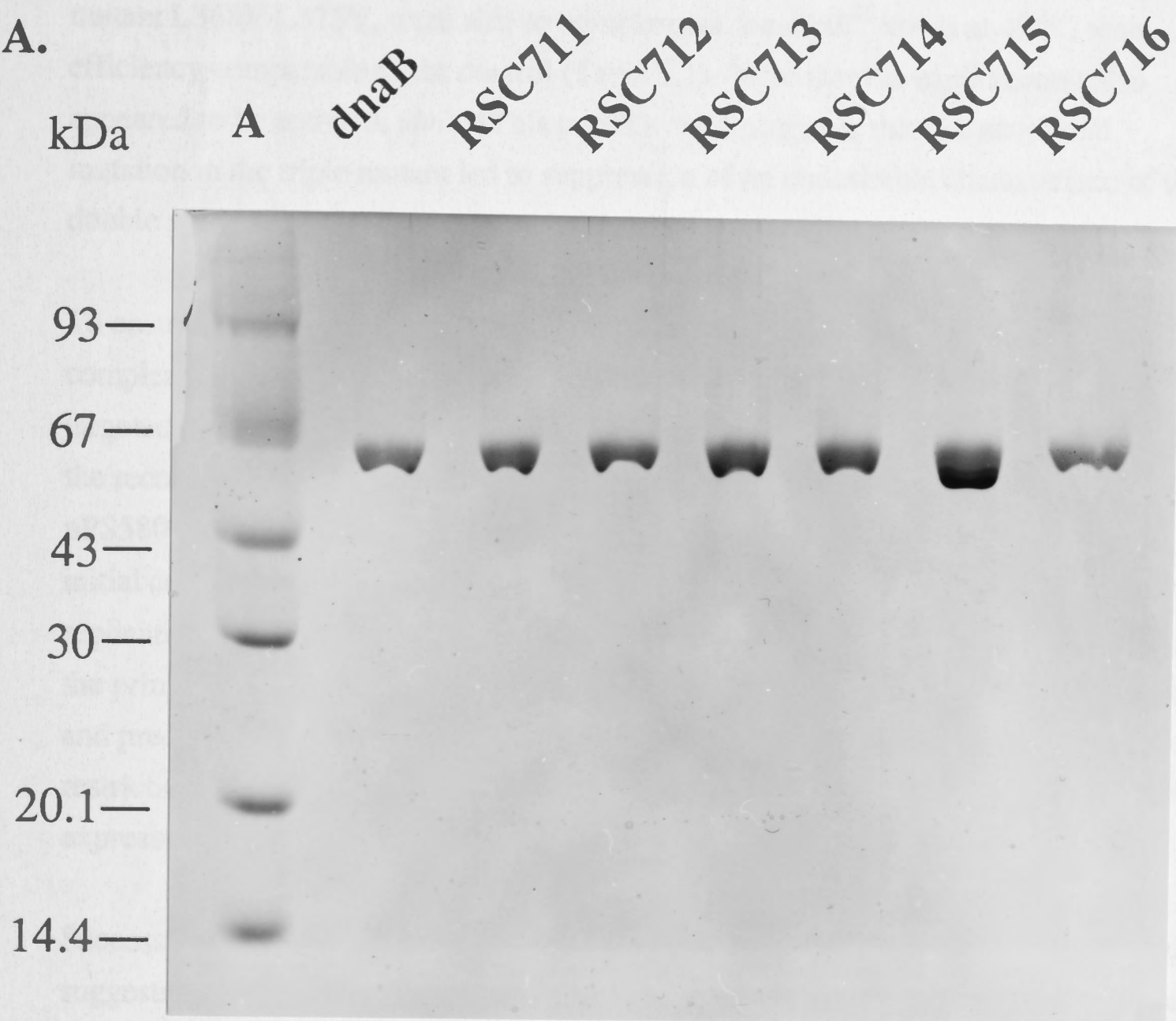
a. Plating efficiency = number of colonies at 42 °C/number of colonies at 30 °C.

A. Mutant	B. Results of Purification		C. M_r (Estimation by gel filtration)	D. Complementation of SG1692 <i>dnaB</i> ^{ts} <i>recA</i>	
	Protein (mg)	$10^{-3} \times$ Specific activity (units/mg)		Strain	Plating ^a efficiency
wild-type <i>dnaB</i>	-----	120	304,000	(RSC807/pPS620) (RSC803/pPS619)	0.74 0.74
L375V (RSC697/pPS572) (RSC711/pPS578)	2.02	115	301,000	(RSC795/pPS612)	0.75
L368V (RSC698/pPS573) (RSC712/pPS579)	1.92	111	279,000	(RSC796/pPS613)	0.69
L368V.L375V (RSC699/pPS574) (RSC713/pPS580)	0.16	88.5	292,000	(RSC797/pPS614) (RSC802/pPS618)	<10 ⁻⁹ <10 ⁻⁹
L382V (RSC700/pPS575) (RSC714/pPS581)	4.62	96.5	276,000	(RSC798/pPS615)	0.74
L368V.L375V.L382V (RSC701/pPS576) (RSC715/pPS582)	0.14	48.5	272000 96,000* 45,000	(RSC799/pPS616)	0.65
L375V.L382V (RSC702/pPS577) (RSC716/pPS583)	2.65	95.3	304,000	(RSC800/pPS617)	0.71

Figure 6.11

Partially purified dnaB* protein from strains RSC711-716 were isolated following DEAE-cellulose anion-exchange chromatography. Mutant dnaB* proteins (1,000 units) were purified as described (Section 6.2.4) and separated by 12-% polyacrylamide gel electrophoresis in the presence of SDS. Protein markers (sizes in kDa) were as in Section 2.4.3 and an authentic sample of dnaB protein (Fraction VI, 1000 units; Table 5.3) is indicated.

A.



hexamer of $M_r \sim 300,000$ (Figure 6.12). The L368V·L375V·L382V mutant eluted in three peaks of different apparent molecular weight, the dominant species being that of $M_r 96,000$. This result suggests that the association of dnaB protomers is directly linked to the presence and maintenance of the leucine heptad repeat. This major species is presumably a dimer of two dnaB* protomers ($M_r 52,000$, as predicted by the molecular weight of the protein determined from the *dnaB* gene sequence; Nakayama *et al.*, 1984b). Species consistent with being a hexameric and monomeric form were also observed, although the ratio of the quantity of each species has not been determined.

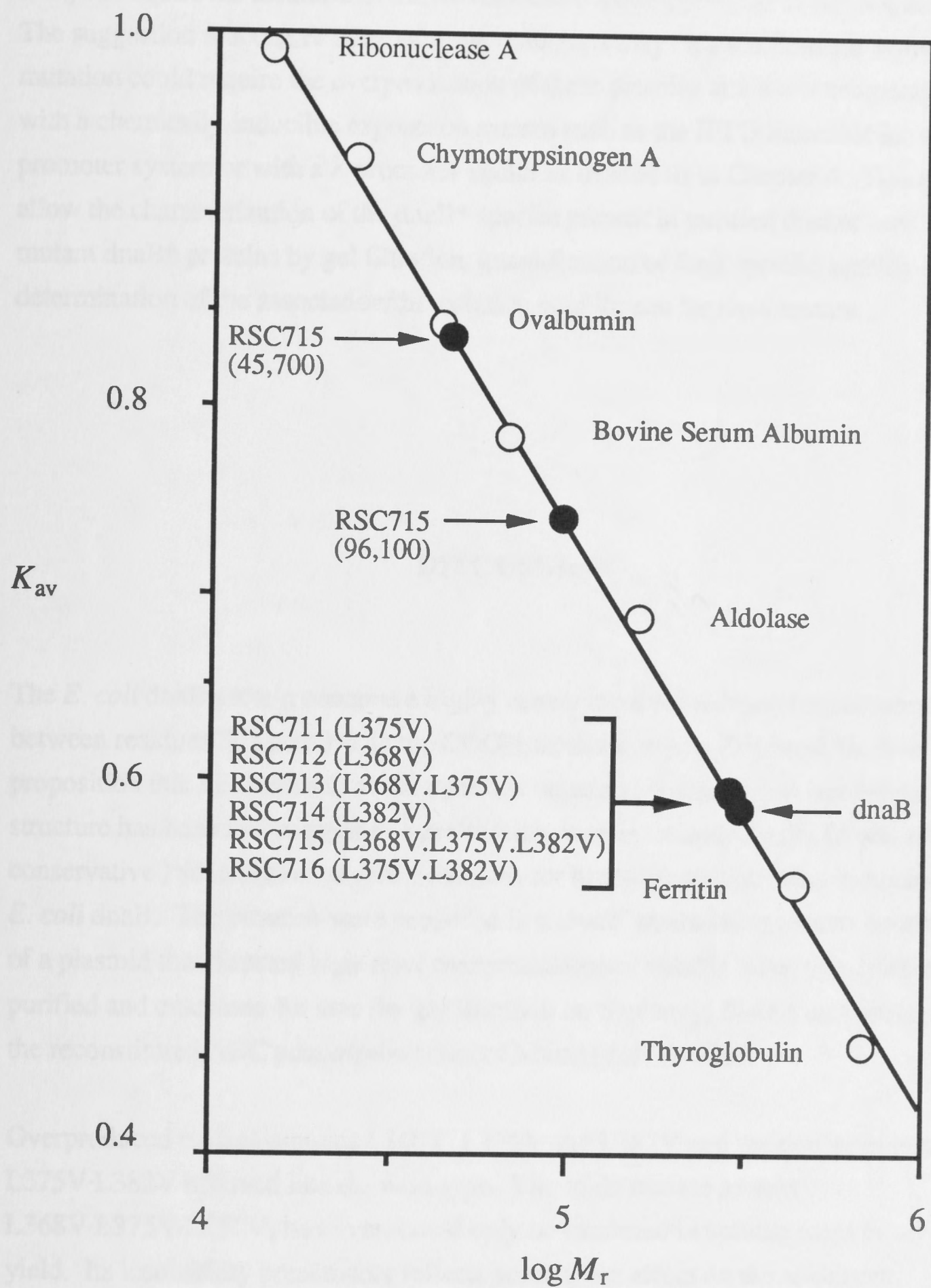
The capacity of the dnaB mutants to promote priming in *E. coli* replication was also examined by complementation of the *dnaB* temperature sensitive strain SG1692. Low-level production of the mutant dnaB* proteins in this strain was achieved by inserting the mutant genes downstream of the *lac* promoter in the cloning vector pMTL22P (Chambers *et al.*, 1988). All mutants, with the exception of the dnaB* mutant L368V·L375V, were able to complement the *dnaB*^{ts} strain at 42 °C with efficiency comparable to the control (Table 6.1). Note that the triple mutant also appeared to be active *in vivo*. [This curious result suggests that the additional mutation in the triple mutant led to suppression of an undesirable characteristic of the double mutant.]

To ensure that the plasmid-borne mutation or mutations responsible for this inability to complement the temperature-sensitive chromosomal mutant is located in the region targeted for site-directed mutagenesis and not a chance mutation elsewhere in the gene, the reconstructed dnaB mutant L368V·L375V with the wild-type 5'-P terminus from pPS580 was also expressed under the *lac* promoter in pMTL23P vector. Like the initial complementation result, this plasmid was also unable to support chromosomal replication in SG1692 at a restrictive temperature. Four other mutants isolated during the primary screening of transformants following site-directed mutagenesis of pPS552 and predicted to contain L368V·L375V alterations in their protein products based on restriction endonuclease digests, gave similar results when cloned in pMTL22P and expressed in SG1692 at 42 °C (results not shown).

Successful complementation in the triple dnaB* mutant protein L368V·L375V·L382V suggests that acquisition of the L382V mutation suppresses the defective phenotype of the L368V·L375V double mutant. The nature of this suppression could be either overcoming a temperature-sensitive characteristic (where the protein is inactivated at a restrictive temperature, such as 42 °C) of the double mutant or reconstitution of a chromosomal replication dependent interaction that is not essential during template

Figure 6.12

Molecular size of purified dnaB* mutant proteins estimated by gel filtration. Proteins were chromatographed on Sephacryl S-400 as described in Section 2.4.5. The K_{av} values $[(V_e - V_o)/(V_i - V_o)]$ determined for each of the purified proteins and the standards (Section 2.4.5) are plotted. The line was determined by linear regression analysis. Due to the low yield of dnaB* from strains RSC713 and RSC715, proteins from these sources could not be detected accurately using a UV monitor attached to the gel filtration column. Portions of the protein in fractions collected during gel filtration from these two sources were run out on a 12-% polyacrylamide gel in the presence of SDS and stained using the silver stain method. The elution volumes (V_e) of the two proteins could then be established by plotting values obtained by densitometric scanning of these gels against the elution volume of fractions. Although five of the proteins had molecular weights similar to that of the native enzyme, one eluted as three separate species consistent with it being a mixture largely of a dnaB dimer and monomer (Table 6.1).



replication in the ABC priming assay.

Further experiments to characterize this motif outside of further extensive mutational analysis require the isolation of the two insoluble mutant proteins in larger quantities. The suggestion that one or more of these mutations may be a temperature-sensitive mutation could require the overproduction of these proteins at a lower temperature with a chemically-inducible expression system such as the IPTG inducible *lac* or *tac* promoter system or with a λ promoter vector as described in Chapter 4. This may allow the characterization of the *dnaB** species present in purified double- and triple-mutant *dnaB** proteins by gel filtration, quantification of their specific activity and determination of the association/dissociation equilibrium for each mutant..

DISCUSSION

The *E. coli* *dnaB* protein contains a highly conserved leucine heptad repeat sequence between residues 361 and 391 in the COOH-terminal region (Figure 6.1). The proposition that this region is at the interface between subunits in the hexameric structure has been investigated in a preliminary way by examining the effects of (very conservative) substitution of valine residues for heptad-repeat leucines in hexameric *E. coli* *dnaB*. The proteins were produced in a *dnaB*⁺ strain using mutant derivatives of a plasmid that directed high-level overproduction of soluble wild-type *dnaB* protein, purified and examined for size (by gel filtration on Sephacryl S-400) and activity (in the reconstituted ABC primosome assay of Masai *et al.*, 1990b).

Overproduced mutant proteins L368V, L375V and L382V and the double mutant L375V·L382V behaved like the wild-type. The triple mutant protein L368V·L375V·L382V, however, could only be extracted in soluble form in very low yield. Its insolubility presumably reflects some gross effect on the tertiary or quaternary structure of the protein, consistently with exposure of hydrophobic regions of the protein to solvent. A small amount of soluble protein was isolated. It was found to have reduced (*ca.* 44 % of wild-type) activity and appeared to be a mixture of oligomeric species, with the dimer predominant. Since the yield was low, it is probable that the soluble form contained some proportion of wild-type subunits expressed from the host chromosome. Although further studies are required to

discount alternative explanations, the data are consistent with the proposition that the mutant protein was unable to form hexamers, with the monomer not amenable to analysis because of its insolubility. These results, in combination with the fact that the carboxy-terminal domain of dnaB protein is known to be responsible for oligomerization of the native protein (Nakayama *et al.*, 1984a), provide evidence that the heptad repeat may be a site for protomer association.

Investigations into the requirement of the leucyl residues in other heptad repeat motifs associated in protein interactions have found the presence of leucine at every seventh position indispensable. The stringency of this requirement is evident in substitution experiments where replacing two or more leucine residues with any amino acid other than proline in a standard four unit repeat leads to dissolution of protein association (Gentz *et al.*, 1989; Turner and Tjian, 1989; Hu *et al.*, 1990). Such a requirement is also evident in the heptad repeat associated with protein oligomerization in dnaB. However, in this example the substitution of three leucines in the repeat was necessary to destabilize interactions involved within the dnaB hexamer.

Unfortunately little is known of the specific roles leucines have in the stabilization of these motifs and the chemical interactions that make them indispensable. However, the ability to promote priming of replication in the ABC-priming assay of each of the purified mutant proteins suggests that these highly conservative leucines in dnaB do not form part of an essential catalytic site and are therefore more likely involved in maintaining protein structure or, as proposed here, associated with protein-protein interactions within the dnaB complex. The possibility that this motif is also involved in interactions with DNA in semblance to the eukaryotic transcriptional activators has not yet been investigated.

As indicated previously, the heptad repeat motif in dnaB departs from the proposed rules delineating potential leucine zippers by maintaining a highly conserved proline at its centre. The calculated probability that the heptad repeat in dnaB is likely to form a coiled coil by interacting with an identical amino acid sequence was high when compared to sequences of known coiled-coil containing proteins (Lupas *et al.*, 1991). If we assume that the heptad repeat in dnaB forms an α -helix then it must form such a structure in the presence of the proline.

Early statistical analyses of protein secondary structures demonstrated that proline residues seldom occur internally in α -helices. Proline is almost locked into a turn by virtue of its side chain cyclization and has been considered a classic 'helix breaker' (Chou and Fasman, 1978). It imparts an unacceptable disruption of the α -helix

hydrogen bonding, and by dictating a kink in the structure it also disturbs the packing of the helix side chains. However, a survey of 291 helices from 57 crystal structures of globular proteins has identified several proline-kinked α -helices (Barlow and Thornton, 1988). These tend to be longer than average and therefore presumably more stable and better able to tolerate the lost hydrogen bonding which stabilizes the α -helical structure.

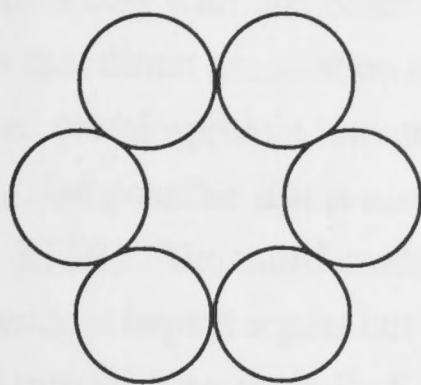
Of all the globular proteins whose crystal structure have been determined and are shown to contain such kinked α -helices, the proline responsible is also conserved in the relevant protein sequences of their known homologues (Barlow and Thornton, 1988). Since the proline residues are conserved in all cases, the question arises as to whether the kinks in these α -helices serve some useful purpose. Certainly for some of the helices there are indications that the kinks may have a definite functional role. The proline (Pro358) in catalase, for example, allows a redirection of the α 9 helix in order to optimise the interaction between Tyr357 and the haem group (Murthy *et al.*, 1981). In conjunction with these observations the highly conserved prolyl residue (Pro378) at the centre of the heptad repeat in dnaB, rather than disrupt the α -helix, may in fact impart some necessary structural variation required for dnaB protomer association. Whereas, in examples to date, leucine zippers facilitate interactions only in dimeric complexes, the dnaB protein and related helicases are hexameric proteins. It seems plausible that bending of a helix at the proline residues might permit six monomer units to pack together in a structural variant of the coiled-coil interaction between leucine zippers. All of these proteins have an additional heptad repeat of hydrophobic amino acids in the g position (in the standard nomenclature for coiled-coils, the zipper leucines are in the d position; *e.g.* Schuermann *et al.*, 1991), resulting in another hydrophobic face to the helix that might promote packing of the hexamers *via* the formation of a large hydrophobic core. To examine the role of the proline residue, a fourth oligonucleotide has been synthesized (RSC41, Figure 6.4) which will direct the substitution of the proline (Pro378) to an alanine residue and will accompany the construction of a *Hind*II restriction endonuclease site in the resultant double-stranded DNA.

In light of this we can draw several models for protomer association within dnaB. The hexameric nature of the protein suggests one of three packing states, either planar or, a trigonal prism or an octahedron (Figure 6.13A), all of which may provide identical environments for each dnaB protomer. The definite polarity of its translocation on single-stranded DNA (LeBowitz and McMacken, 1986) may suggest a certain degree of asymmetry in the protein.

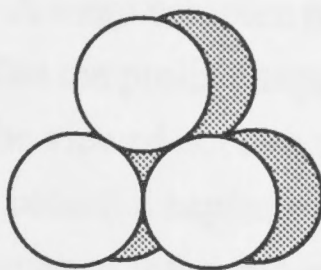
Figure 6.13

Models for the association of dnaB protomers. The hexameric nature of the dnaB protein dictates that three packing states are possible in which each dnaB protomer shares an identical environment (A). In all three models it is possible that the structure is maintained through the association of three dimers each of which is formed through the interactions of their heptad repeats. The repeating heptad between residues 361 and 392 of the *E. coli* dnaB amino-acid sequence suggests an α -helical structure, consistent with predictions of secondary structure (Nakayama *et al.*, 1984a). However, a proline at the centre of this heptad repeat indicates that there may be a kink, with an angle θ , in the amino-acid backbone (B). Such a kink might create two structurally separate heptad repeats which are now able to interact on an individual basis with heptad repeat motifs in other subunits (C). In addition to interactions producing dimeric units, this may involve (i) the successive interaction of heptad repeats from separate dnaB protomers forming a planar hexameric structure, or (ii) the similar formation of trimers which would then require the presence of other protein-protein interactions to form an octahedral or trigonal prismatic structure. Other proline-kinked α -helix interactions are predicted to form parallel coiled-coils with hydrophobic contribution from either the a or g residues as opposed to the original proposal for leucine zippers involving the antiparallel association of two heptad repeats allowing the sequential interdigitation of leucyl side chains. The dnaB protein maintains a high proportion of hydrophobic residues in the a position which could enable the second-form of coiled-coil interaction or, in a more complex case, increase the strength of interactions between three heptad repeats (D). Such three-stranded coiled coil structures which form molecular trimers have been previously identified (Cohen and Parry, 1986).

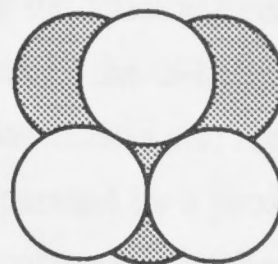
A) Models for the structure of the dnaB protein.



i) Planar Hexagonal

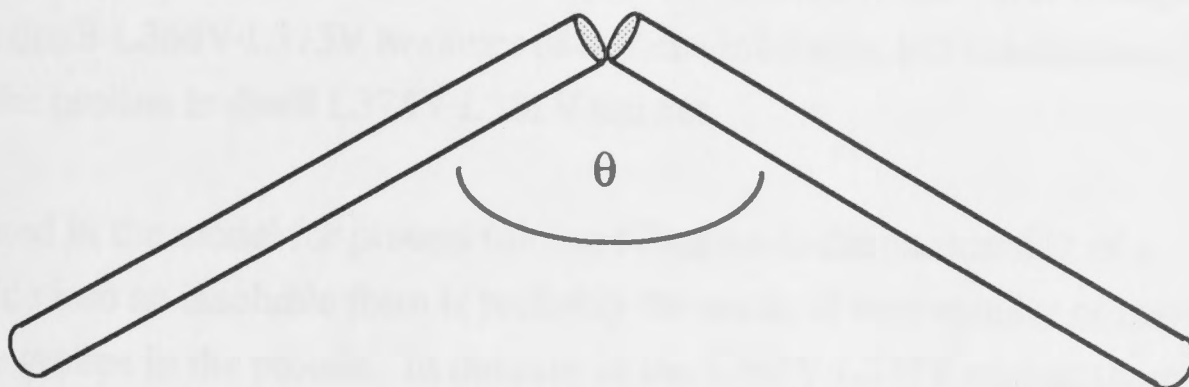


ii) Trigonal prism



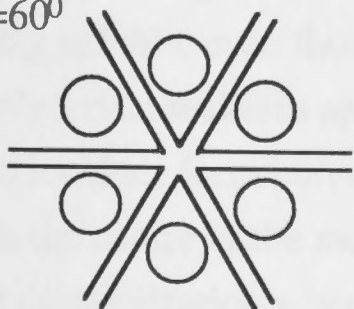
iii) Octahedral

B) A proline-kinked α -helix.



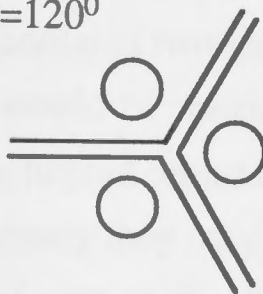
C) Possible interactions involved between kinked leucine zippers.

i) $\theta=60^\circ$



6-fold interaction

ii) $\theta=120^\circ$



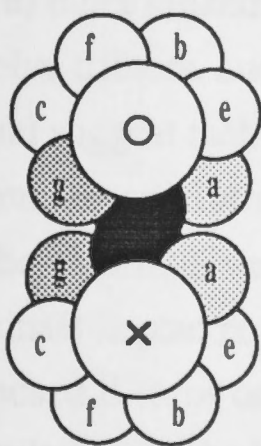
3-fold interaction

iii) $\theta=180^\circ$

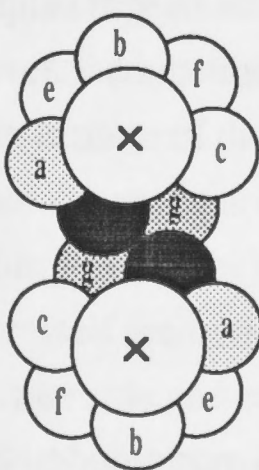


2-fold interaction

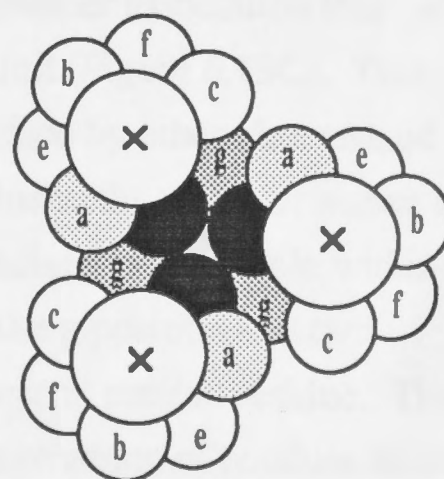
D) End-view depiction of α -helices and their leucine zipper interactions.



i) Antiparallel association



ii) Parallel association



iii) Trimeric association

Hexamer formation in these models is possible if each four unit heptad repeat interacted solely with one other from a separate protomer of dnaB. This automatically assumes that dimer association between dimers will be directed by other as yet undefined protein-protein interactions. A more complex model for the leucine heptad repeat is also possible if it is assumed that the proline imparts a kink in the α -helix (Figure 6.13B). The motif could then be viewed not as a single interaction site for one other identical heptad repeat but as two potential heptad repeats separated by a proline induced turn with an angle θ . Such a series of interactions may in part explain the observed phenotypic differences between the two dnaB* mutants L368V·L375V (insoluble) and L375V·L382V (soluble). The substitution of two leucyl residues in the dnaB protein both NH₂-terminal to the proline residue (Pro378), is enough to cause the dnaB L368V·L375V hexamer to become insoluble, but substitutions that straddle the proline in dnaB L375V·L382V are not.

As proposed in the model for protein folding (Chapter 4) the partitioning of a polypeptide into an insoluble form is probably the result of accessibility of solvent to non-polar groups in the protein. In the case of the L368V·L375V mutant there may be enough localized destabilization between the two heptad repeats of separate protomers to cause a partial abrogation of the interaction allowing the hydrophobic residues within the heptad repeat to become exposed to the solvent. The other double mutant is completely soluble, even though it too contains two leucine substitutions in the heptad repeat only seven residues apart. This would be consistent with the proposal that these two leucines form part of separate heptad repeats isolated by the presence of the proline in the centre of the motif. Separately they may be incapable of creating enough localized destabilization when associated with an identical mutant protomer to affect their heptad repeat interactions.

A situation where a central proline creates a kink enables interactions with two (or more) other kinked heptad repeats allowing models for protomer association that involve trimers, and even the hexamer as an asymmetric unit (Figure 6.13C). This would suggest that stabilization of the protein-protein interface by other electrostatic interactions and structural features are probably required due to the short nature of the heptads involved (only two units). Two features immediately observable within the dnaB heptad repeat itself are hydrophobic residues in the g position and two proposed β -turns on either side and equidistant from the central proline residue. The contribution to the hydrophobic core of the heptad repeat interaction of residues at the g position as well as the a position may well explain the great stability within the dnaB hexamer (*e.g.* see Figure 6.13D). The β -turns predicted from the primary amino-acid sequence (Nakayama *et al.*, 1984b) are centred at residues 366 and 387 in the *E. coli*

dnaB sequence and may also add secondary structure that could further stabilize associations between protomers.

The identification of the heptad repeat associated with protein oligomerization in dnaB was dependent on alignment of sequences of several analogous proteins from other sources (Appendix B). Comparisons have also been carried out with the bacteriophage T7 gene 4 product, a primase-helicase, which shares considerable sequence identity with the COOH-terminal domain of other dnaB-like proteins (Wong *et al.*, 1988). Inspection of the amino-acid sequence reveals that this protein maintains the great majority of those residues conserved within the sequences of the five other dnaB-like proteins within the region of the heptad repeat (Figure 6.14), but has only one of the four conserved leucine residues identified in other dnaB proteins and has no central proline residue in this region. The T7 gene 4 products are co-purified as 56-kDa and 63-kDa components (Hinkel and Richardson, 1975) that display helicase and helicase-primase activities, respectively (Nakai and Richardson, 1988; Bernstein and Richardson, 1989). The fact that these two proteins can also be isolated separately suggests that the small form of the protein is in effect a natural but functional dnaB leucine-heptad repeat mutant in which the protein no longer requires subunit assembly for functional replication of the T7 viral genome. Apart from strengthening the argument that sequences necessary for dnaB protomer association are probably located within this heptad repeat, this finding has several interesting implications for DNA replication.

The current view of T7 DNA replication involves initiation at the primary origin of replication through transcription by T7 RNA polymerase from one of the promoters within the origin sequence (Richardson *et al.*, 1987). Displacement of the T7 RNA polymerase by T7 DNA polymerase results in a DNA polymerase complex tightly bound to the 3'-OH end of the transcript in precisely the conformation needed to catalyse DNA synthesis. However, this complex cannot polymerize nucleotides through duplex DNA without the prior association of a monomer of T7 gene 4 helicase-primase which translocates unidirectionally in the 5'→3' direction unwinding the duplex at the replication fork. The primase-helicase also catalyses the production of tetranucleotide template-specified primers on the lagging strand which can then be extended into full length Okazaki fragments by the T7 DNA polymerase (Figure 6.15).

This model is consistent with the notion that an asymmetric arrangement of protomers within the dnaB hexamer is not required for the specificity of its 5'→3' direction of translocation but is intrinsic to the structure of the individual dnaB protomers. It

Figure 6.14

Alignment of the residues of the bacteriophage T7 gene 4 protein with the conserved residues within the leucine heptad repeat identified in *E. coli* dnaB and four homologous proteins. The consensus sequence represents those amino acids conserved between all five dnaB analogues (Appendix B) displayed as the *E. coli* sequence. Similarities are indicated by boxed amino acids: hatched = same amino acid; open = amino acids belonging to the same group (hydroxyl/small aliphatic: A, G, S, T; acid and acid amide: N, D, E, Q; basic: H, R, K; aliphatic: M, I, L, V; or aromatic: F, Y, W). The second sequence displays an homologous segment of polypeptide identified within the bacteriophage T7 gene 4 protein (Dunn and Studier, 1983; Appendix B). Only similarities between this sequence and the consensus sequence are indicated: hatched = amino acids consistent with the consensus sequence; open = amino acids belonging to the same group as the consensus sequence where the consensus sequence is absolutely conserved. Arrows indicate the positions of those leucine residues predicted to participate in protomer association in the *E. coli* dnaB protein and the asterisk indicates the highly conserved proline residue at the centre of the motif. Amino-acid abbreviations are the standard one letter abbreviations (see Appendix B).

340
 |
 Consensus G I G L I M I D Y L Q L M - R V P A L S D N R T L E I A E I
 T7 gene 4 G C D V I I L D H I S I V V S A S G E S D E R - K M I D N L
 350
 |
 360
 | ↓

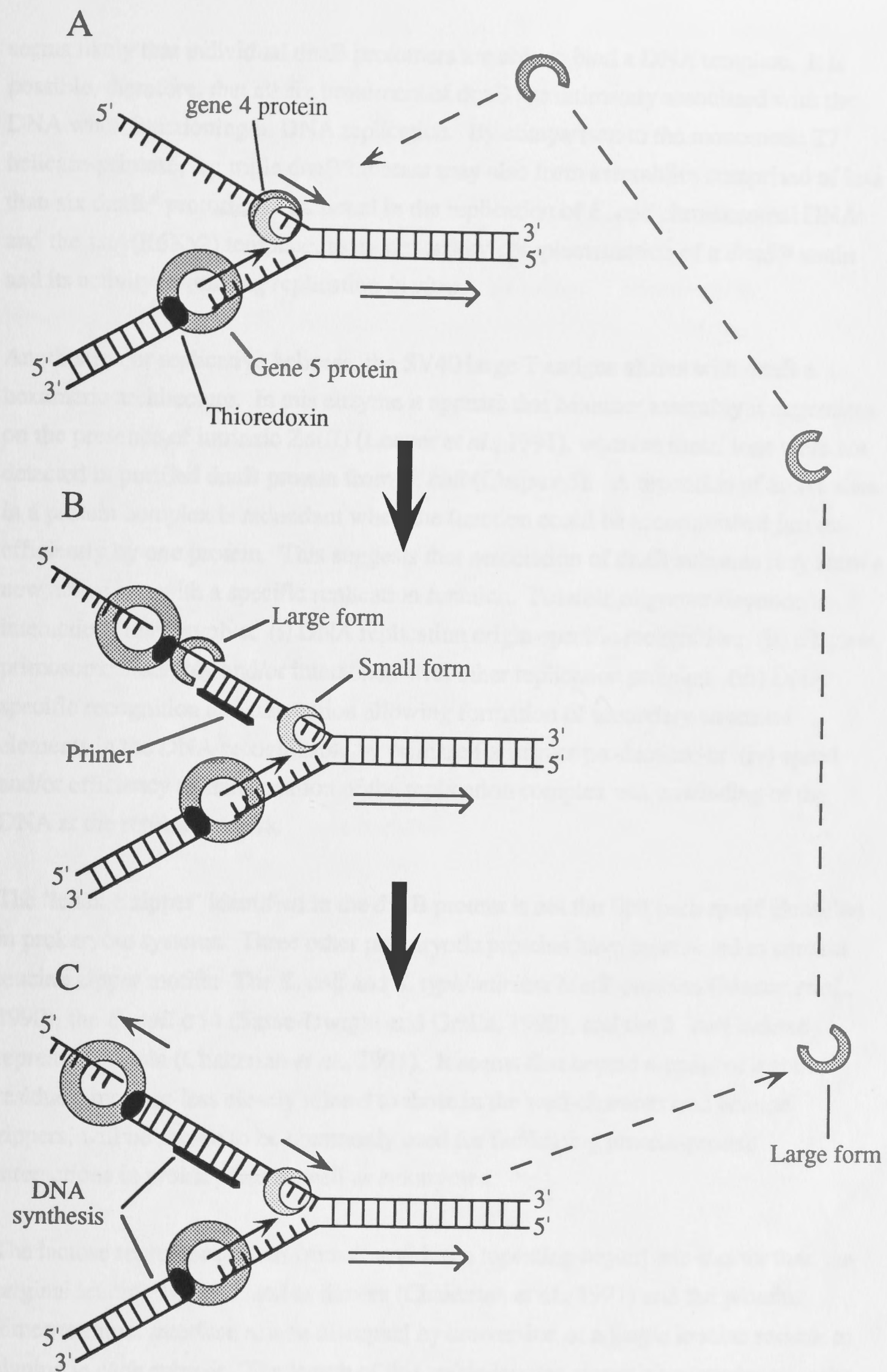
↓ 370
 |
 Consensus S R S L K A L A K E L N V P V V A L S Q L N R S L E Q R A -
 T7 gene 4 M T K L K G F A K S T G V V L V V I C H L K N P D K G K A H
 ↓ * 380
 |
 ↓ 390
 |

400
 |
 Consensus - D K R P V - N S D L R E S G S I E Q D A D L I M F I Y R D
 T7 gene 4 E E G R P V S I T D L R G S G A L R Q L S D T I I A L E R N
 410
 |
 420
 |

Figure 6.15

A model for leading and lagging strand synthesis at the replication fork of bacteriophage T7. A. The small form of the gene 4 protein serves as a helicase for leading strand synthesis by T7 DNA polymerase (gene 5 protein-thioredoxin complex). The large form of gene 4 protein binds to the small form to create an active primase. B. As the small form-large form complex encounters a primase recognition site, the large form binds at the site, dissociating from the small form, and catalyses the synthesis of a tetranucleotide primer. The small form continues to serve as helicase for leading strand synthesis. Its movement $5' \rightarrow 3'$ along single-stranded DNA is uninterrupted by primer synthesis. C. As DNA synthesis catalysed by T7 DNA polymerase extends the primer, the large form dissociates from the template and is available to catalyse primer synthesis again at the replication fork. Arrows indicate the direction protein complexes move along the DNA template and open arrows indicate the direction of fork movement.

Figure adapted from Nakai and Richardson (1988).



seems likely that individual dnaB protomers are able to bind a DNA template. It is possible, therefore, that all six protomers of dnaB are intimately associated with the DNA when functioning in DNA replication. By comparison to the monomeric T7 helicase-primase, the triple dnaB* mutant may also form assemblies comprised of less than six dnaB* protomers functional in the replication of *E. coli* chromosomal DNA and the *ssiA*(R6K γ 2) template, as evidenced by complementation of a *dnaB*^{ts} strain and its activity in priming replication *in vitro*.

Another major replicative helicase, the SV40 large T antigen shares with dnaB a hexameric architecture. In this enzyme it appears that hexamer assembly is dependent on the presence of intrinsic Zn(II) (Loeber *et al.*, 1991), whereas metal ions were not detected in purified dnaB protein from *E. coli* (Chapter 5). A repetition of active sites in a protein complex is redundant when the function could be accomplished just as efficiently by one protein. This suggests that association of dnaB subunits may form a new active site with a specific replication function. Possible oligomer-dependent interactions may involve: (i) DNA replication origin-specific recognition; (ii) efficient primosome assembly and/or interaction with other replication proteins; (iii) DNA specific recognition and interaction allowing formation of secondary structural elements in the DNA recognizable by primase for primer production; or (iv) speed and/or efficiency of translocation of the replication complex and unwinding of the DNA at the replication fork.

The 'leucine zipper' identified in the dnaB protein is not the first such motif identified in prokaryote systems. Three other prokaryotic proteins have been noted to contain leucine zipper motifs: The *E. coli* and *S. typhimurium* MetR proteins (Maxon *et al.*, 1990), the *E. coli* σ 54 (Sasse-Dwight and Gralla, 1990), and the *E. coli* lactose repressor protein (Chakerian *et al.*, 1991). It seems that heptad repeats of leucine residues, more-or-less closely related to those in the well-characterized leucine zippers, will be shown to be commonly used for facilitating protein-protein interactions in prokaryotes as well as eukaryotes.

The lactose repressor protein from *E. coli* has a repeating heptad one shorter than the original leucine zipper found in dimers (Chakerian *et al.*, 1991) and the proteins dimer-tetramer interface can be disrupted by conversion of a single leucine residue to alanine in each subunit. The length of this stable leucine zipper gives credence to the proposal that dnaB may form a leucine zipper interface of only two heptad repeat units which are further stabilized by several other interactions.

In drawing models of interactions likely to occur in the hexamerization of dnaB, two

major assumptions have been made. First, that the heptad repeat in dnaB actually forms an α -helix and secondly that the proline residue contributes substantially to the secondary structure of the motif and therefore, to the type of interactions possible with other dnaB protomers. These assumptions have been made with a lack of any physical evidence, and it is dangerous to draw conclusions about protein structure on this basis. Only the structure of the protein, determined by X-ray crystallography, will give a clearer view of the type of association present. Whatever the result, the complexity of interactions involved with protomer association in dnaB are not entirely unique. Like all 'leucine zippers', these are more likely a modified variation of the ubiquitous coiled-coil.

The sequence of *priC*, the last primosomal gene to be identified, has recently been reported (Zavitz *et al.*, 1991). This gene encodes the 11.4-kDa PriC protein (protein n") of *E. coli* and like dnaB, displays a brief 3-unit leucine-heptad repeat at its COOH-terminus. It has been suggested, on the basis of its co-purification with dnaB protein in dnaB preparations, that PriC may interact with dnaB through its similar heptad array (Zavitz *et al.*, 1991). If the leucine heptad repeat of dnaB is open to the solvent to allow an interaction with PriC, then presumably the leucine residues play little role in the structure and stability of the protein. This hypothesis is in contradiction to the observed effects substitution of valine for leucine has on the solubility of the dnaB protein. These results suggest that the leucine heptad repeat in dnaB is buried either in each individual subunit or within the dnaB hexamer and therefore unable to partake in surface interactions with other proteins.

It is possible that the interaction of dnaC protein with dnaB is able to mask the leucine heptad repeat from solvent when not in contact with PriC. However, this does not explain the insolubility of dnaB* mutants when overproduced in the presence of dnaC, or the fact that soluble dnaB protein can be isolated separate from other proteins. Because the dnaB* triple mutant is able to complement dnaB functions *in vivo*, one would assume that this mutant is capable of performing efficient lagging-strand replication as part of the fully assembled primosome. To investigate this hypothesis and determine whether the leucine heptad repeat is required for efficient primosomal assembly and the discontinuous initiations of DNA synthesis, the dnaB* mutants described here will in future be assayed in the ϕ X174 SS \rightarrow RF assay.

CHAPTER 7

Domain Structure of the dnaB Protein

Evidence from secondary structural analysis of the primary amino acid sequence (Nakayama *et al.*, 1984b) and partial tryptic hydrolysis of the purified dnaB protein (Nakayama *et al.*, 1984a) have led to the proposal that the protein possesses at least two distinct structural domains (see Figure 6.2). It has also been suggested that dnaB protein contains distinct sites within these domains that promote the various separate activities of the protein, including: (i) various ribo- and deoxyribonucleotide binding activities (Arai and Kornberg, 1981c, d); (ii) DNA-dependent ribonucleotide triphosphatase and dATPase activity (Arai and Kornberg 1981c; Biswas *et al.*, 1986); (iii) single- and double-stranded DNA-binding activity (Arai and Kornberg, 1981c); and (iv) interactions with dnaC (Kobori and Kornberg, 1982c; Wahle *et al.*, 1989a), dnaG (Arai and Kornberg, 1981e) and λ P proteins (McMacken *et al.*, 1983; Biswas and Biswas, 1987).

The COOH-terminal domain has already been shown to contain sites necessary for DNA-dependent ATPase activity (Nakayama *et al.*, 1984a; Biswas and Biswas, 1987) and protein oligomerization (Nakayama *et al.*, 1984a). Evidence in Chapter 6 supports the notion that the latter involves a highly-conserved leucine heptad repeat within this domain. The NH₂-terminal domain, on the other hand, has no independent activity of the dnaB protein, but is implicated in interactions with dnaC (Nakayama *et al.*, 1984a) and possibly other primosomal and replication proteins.

Following the successful purification of the dnaB protein in quantities sufficient for structural studies and active-site chemistry (Chapter 5), it was thought appropriate to initiate a detailed examination of these domains to facilitate and corroborate results of work with the native protein. The two strategies undertaken involved: (i) isolating the separate domains of the protein by manipulating the *dnaB* gene; and (ii) examining a dnaB homologue.

The determination of the terminal amino acids of dnaB fragments created following partial tryptic hydrolysis has identified the termini of potential domains of the protein (Nakayama *et al.*, 1984a, b) and thus their coding regions within the *dnaB* gene. Using recombinant DNA technology, it would be possible to reconstruct the major tryptic peptides at the level of the gene and overproduce the genetically-engineered fragments. This would enable independent characterization of these domains, and assist in determination of the structure of the native protein. The functional and structural analysis of two homologous proteins may also facilitate identification of active-site amino acids and secondary structural features. Several homologues of dnaB protein have been identified (Appendix B) and one of these, the bacteriophage *Pl* *ban* protein, can completely replace *E. coli* dnaB function *in vivo* (Sclafani and

Wechisler, 1981a). The *dnaB* analogue from *S. typhimurium* was of little interest because of its >98 % amino acid homology with *E. coli* *dnaB* (Wong *et al.*, 1988). Of the other *dnaB* analogues, only a *dnaB*-like gene from a *Chlamydia trachomatis* plasmid was immediately available. Studies were therefore directed at the overproduction of the protein product of the gene from *Chlamydia*.

MATERIALS AND METHODS

7.2.1 Plasmids Directing the Overproduction of the COOH-terminus of *dnaB*

7.2.1.1 Manipulation of the *Bgl*III site in the *dnaB* gene

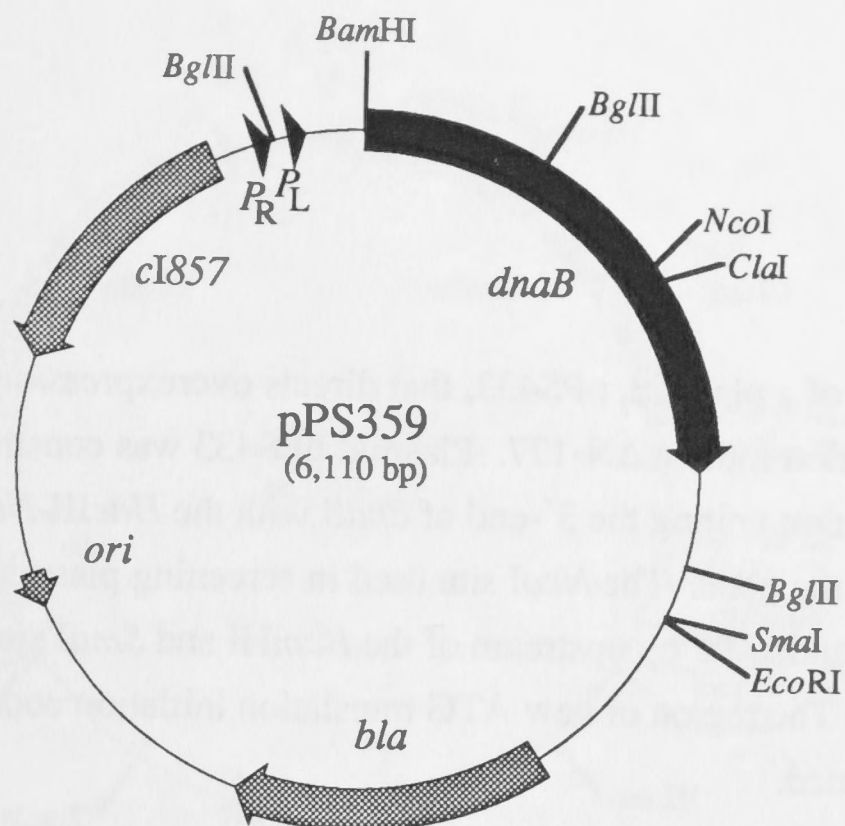
A 1260-bp *Bgl*III DNA fragment identical to that in pPS359 (Section 4.2.3; Figure 7.1) containing the 3'-OH terminus of the *dnaB* gene was end-filled and isolated. This fragment was then ligated into 5'-dephosphorylated *Nco*I-linearized pND217 (Figure 2.1) whose ends had been repaired by treatment with Klenow fragment. Ampicillin resistant transformants of strain AN1459 were selected at 30 °C and screened for temperature sensitivity at 42 °C. Following this, the approximate size of the product plasmids (~5.3 kb) and orientation of the inserted fragments were determined by *Bam*HI, *Nco*I and *Cla*I/*Eco*RI restriction endonuclease digests of plasmid DNA isolated from small-scale analytical preparations. A strain (RSC535) carrying the newly selected plasmid pPS431 was then analysed for the expression of the mutant gene at 42 °C as visualized on SDS-PAGE (Figure 7.4).

7.2.1.2 Manipulation of the *Hae*III sites in the *dnaB* gene

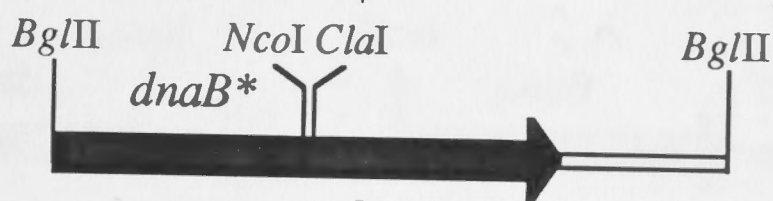
A 369-bp *Hae*III/*Nco*I DNA fragment containing a portion of the 5'-P terminus of the *dnaB* gene and a 1203-bp *Nco*I/*Eco*RI DNA fragment containing the 3'-OH terminus of the *dnaB* gene were isolated from pPS359 (Figure 7.2). These fragments were then ligated in a three-fragment ligation with the *Nco*I-linearized and end-filled plasmid vector pND217 which had been further digested with *Eco*RI. Ampicillin resistant transformants of strain AN1459 were selected at 30 °C, and were screened first for plasmids of the anticipated size (~5.7 kb) and, secondly, for the successful reconstruction of the 3'-OH terminus of the *dnaB* gene by *Sma*I/*Nco*I and

Figure 7.1

Construction of a plasmid, pPS431, that directs overexpression of the *dnaB* NH₂-terminal deletion mutant Δ N-156. Plasmid pPS431 was constructed by ligating a blunt-ended *Bgl*III fragment, containing the 3' end of *dnaB*, to a new ATG translation initiation codon from *Nco*I linearized pND217 whose ends were repaired with Klenow fragment. The *Nco*I and *Cla*I sites used in screening of plasmids are centred 434 bp downstream of the *Bgl*III site in the coding region of *dnaB* and 816 bp upstream of the *Eco*RI site of the vector, respectively. The sequence of the region immediately surrounding the newly acquired ATG translation initiation codon which starts the truncated *dnaB* gene is given.



1. *Bgl*III
2. Fragment isolation



3. *Nco*I-linearized and end-filled
pND217 + ligase

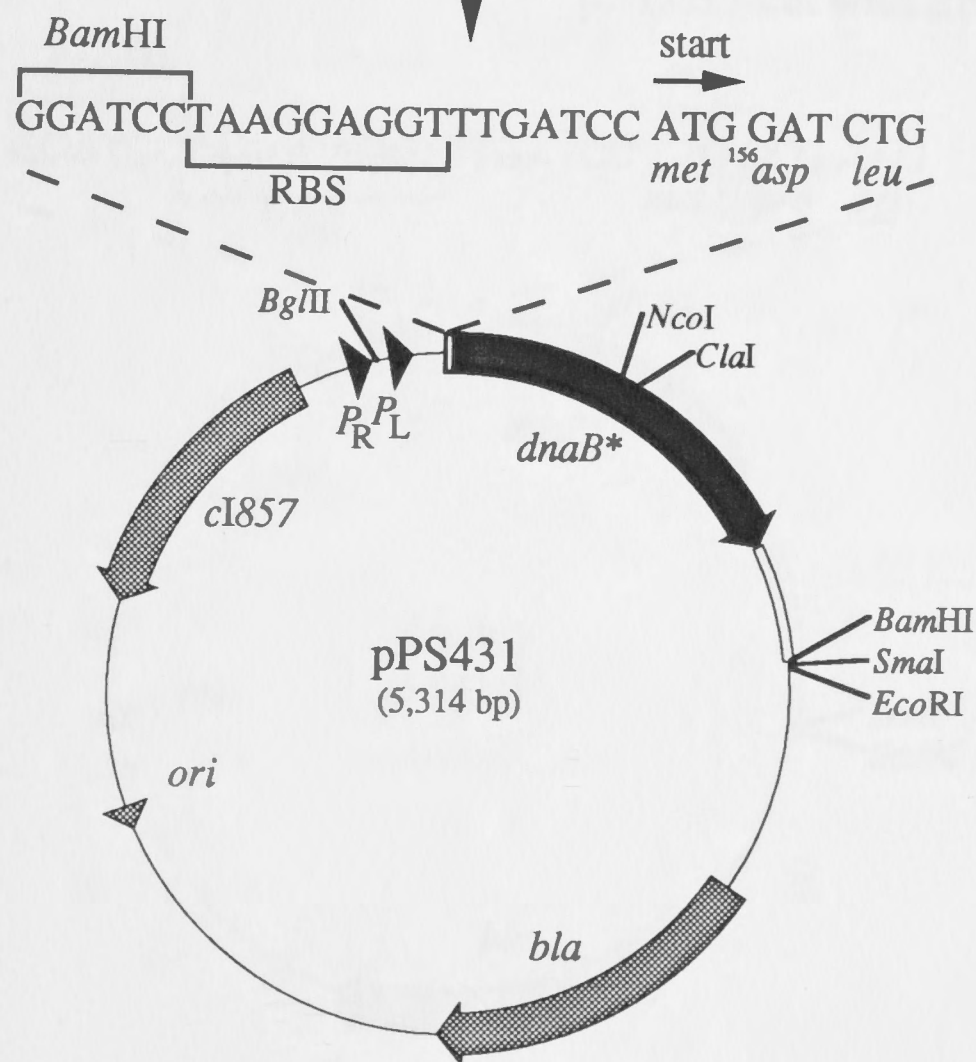


Figure 7.2

Construction of a plasmid, pPS433, that directs overexpression of the *dnaB* NH₂-terminal deletion mutant Δ N-177. Plasmid pPS433 was constructed using a 3-way fragment ligation uniting the 3'-end of *dnaB* with the *Hae*III-*Nco*I fragment isolated from the 5' end of the gene. The *Nco*I site used in screening plasmids is centred 279 bp downstream and 1198 bp upstream of the *Bam*HI and *Sma*I sites of the vector, respectively. The region of new ATG translation initiation codon of the truncated *dnaB* gene is indicated.

*Bam*HI/*Nco*I digests of plasmid DNA isolated from small-scale analytical preparations. A strain (pPS537) carrying the newly selected plasmid, pPS433, was then analysed for the expression of the mutant gene at 42 °C as visualized on SDS-PAGE (Figure 7.4).

7.2.2 Plasmids Directing the Overproduction of NH₂-terminal Regions of *dnaB*

The plasmid pPS359 (Section 4.2.3) was linearized by a partial *Hinf*I digest and the linear plasmid was isolated following repair of the ends by treatment with Klenow fragment. The plasmid was then recircularized by intramolecular blunt-end religation (Figure 7.3). Ampicillin resistant transformants of AN1459 were selected at 30 °C and plasmid DNA isolated from small-scale analytical preparations was screened for the anticipated size (~6.1 kb). These plasmids were further screened by comparing fragment patterns between plasmid DNA and pPS359 digested with *Hinf*I, revealed by agarose gel electrophoresis. Those plasmids without *Hinf*I fragments corresponding to *Hinf*I restriction endonuclease sites within the *dnaB* gene were then selected and analysed for the expression of a mutant *dnaB* gene at 42 °C as visualized on SDS-PAGE (Figure 7.4). Several *dnaB* mutants were isolated, including two which produced COOH-terminal deletion mutants of the *dnaB* protein. These included the insertion of ochre stop codons immediately following the amino acid residues Glu182 (pPS503) and Glu404 (pPS501) and a amino acid insertion mutant containing a tyrosine residue between residues 314(Asp) and 315(Ser) of wild-type *dnaB* (pPS502).

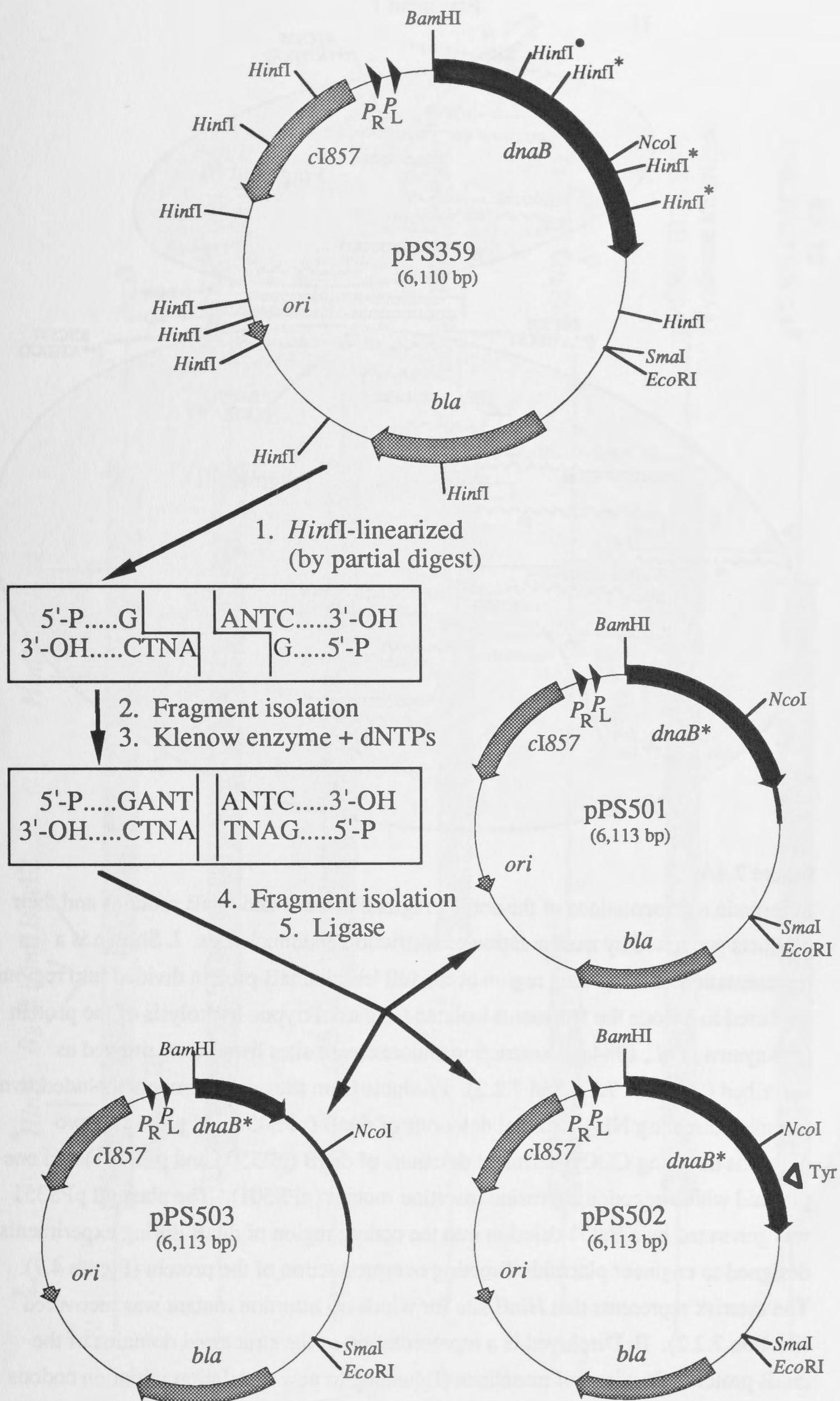
7.2.3 A Plasmid Designed for the Overproduction of "dnaB" from *C. trachomatis*

A plasmid constructed by insertion of a 2277-bp *Sma*I/*Eco*RI DNA fragment containing *ORF3* from the *C. trachomatis* plasmid pCTT1 into pUC18 was obtained from Dr K.S. Sriprakash (Menzies School of Health Research, Darwin, N.T.). From this a 1685-bp *Hpa*I/*Eco*RI fragment containing all of the *ORF3* and 36 bp of DNA upstream of the ATG translation initiation codon was isolated and inserted between the *Hpa*I and *Eco*RI sites of pPT150 (Figure 2.1). Ampicillin resistant transformants of AN1459 were selected at 30 °C and the new plasmid (pPS439) containing the *ORF3* was obtained following screening by *Hpa*I and *Eco*RI restriction endonuclease digestions of plasmid DNA isolated from small-scale analytical preparations. Plasmid

Figure 7.3

Plasmids directing the overproduction of two COOH-terminal deletions of *dnaB* and a *dnaB* tyrosine insertion mutant. The plasmid pPS359 (Section 4.2.3) was linearized by a partial *HinfI* digest, end-filled and recircularized. The insertion mutants that were produced following this procedure were screened by *HinfI* digest and observing fragment patterns compared to pPS359/*HinfI*. Three plasmids were isolated from this exercise encoding insertion mutants at those *HinfI* sites marked with an asterisk. These were shown to have produced two new 3' deletion mutants of *dnaB* by insertion of an ochre stop codon (pPS503, pPS501) and a tyrosine insertion (pPS502). A fourth insertion mutant at the other *HinfI* site in *dnaB* was not isolated. The region of the mutation and the predicted newly engineered coding region of the COOH-terminal deletions are given below together with the relevant restrictive endonuclease digest pattern expected for these mutations.

		<i>HinfI</i> fragments	
		lost	gained
	<i>leu ala glu ***</i>	457	
pPS503	5'.....CTG GCT GAA TAA TCCCGCGTC.....3'	+ →	677
		220	
	<i>leu arg glu ***</i>	269	
pPS501	5'.....CTG CGT GAA TAA TCTGGCGTC.....3'	+ →	766
		497	
	<i>ile asp asp tyr pro pro ala</i>	269	
pPS502	5'.....ATC GAT GAC TAC CCT CCG GCC...3'	+ →	726
		457	



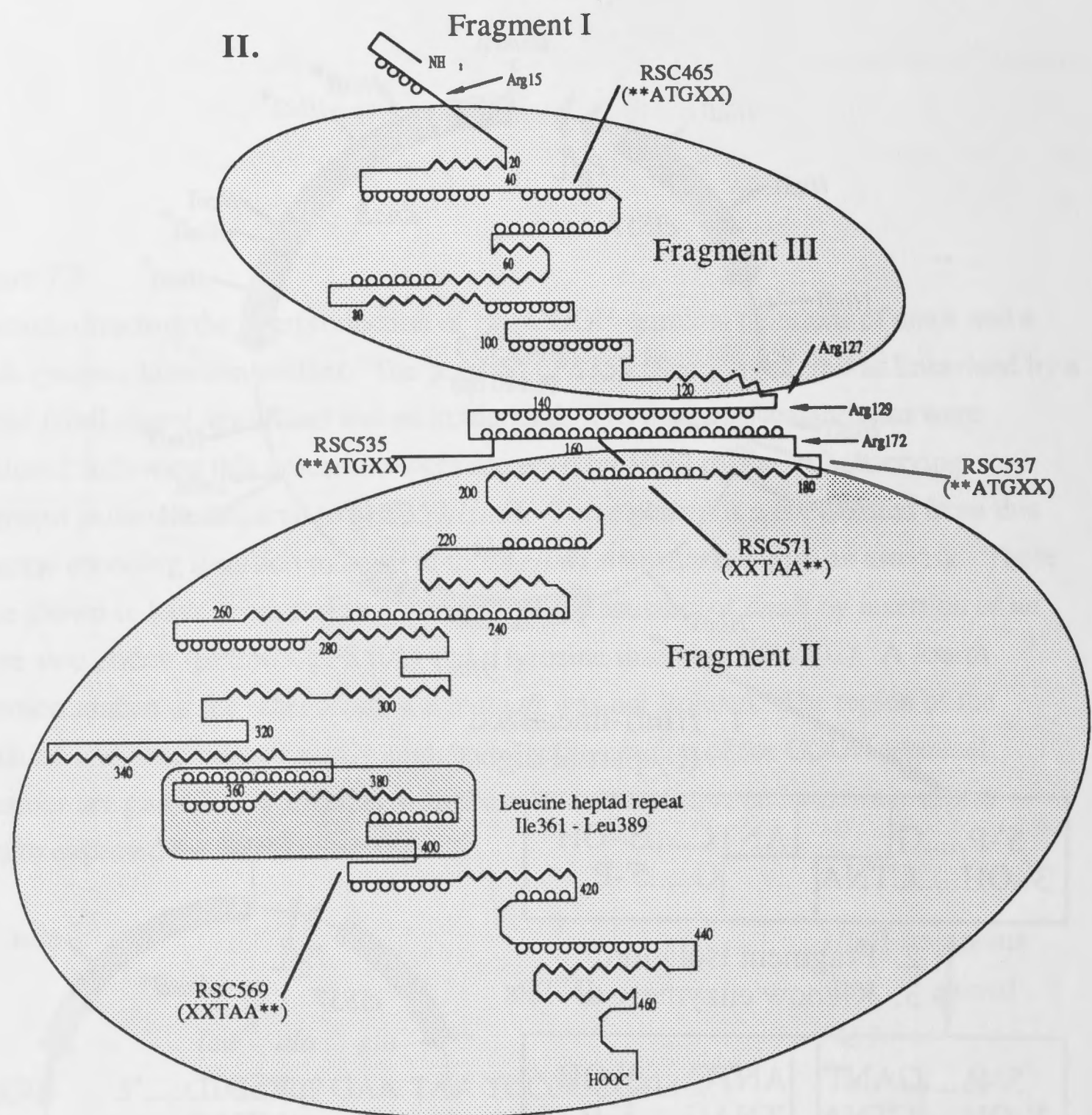
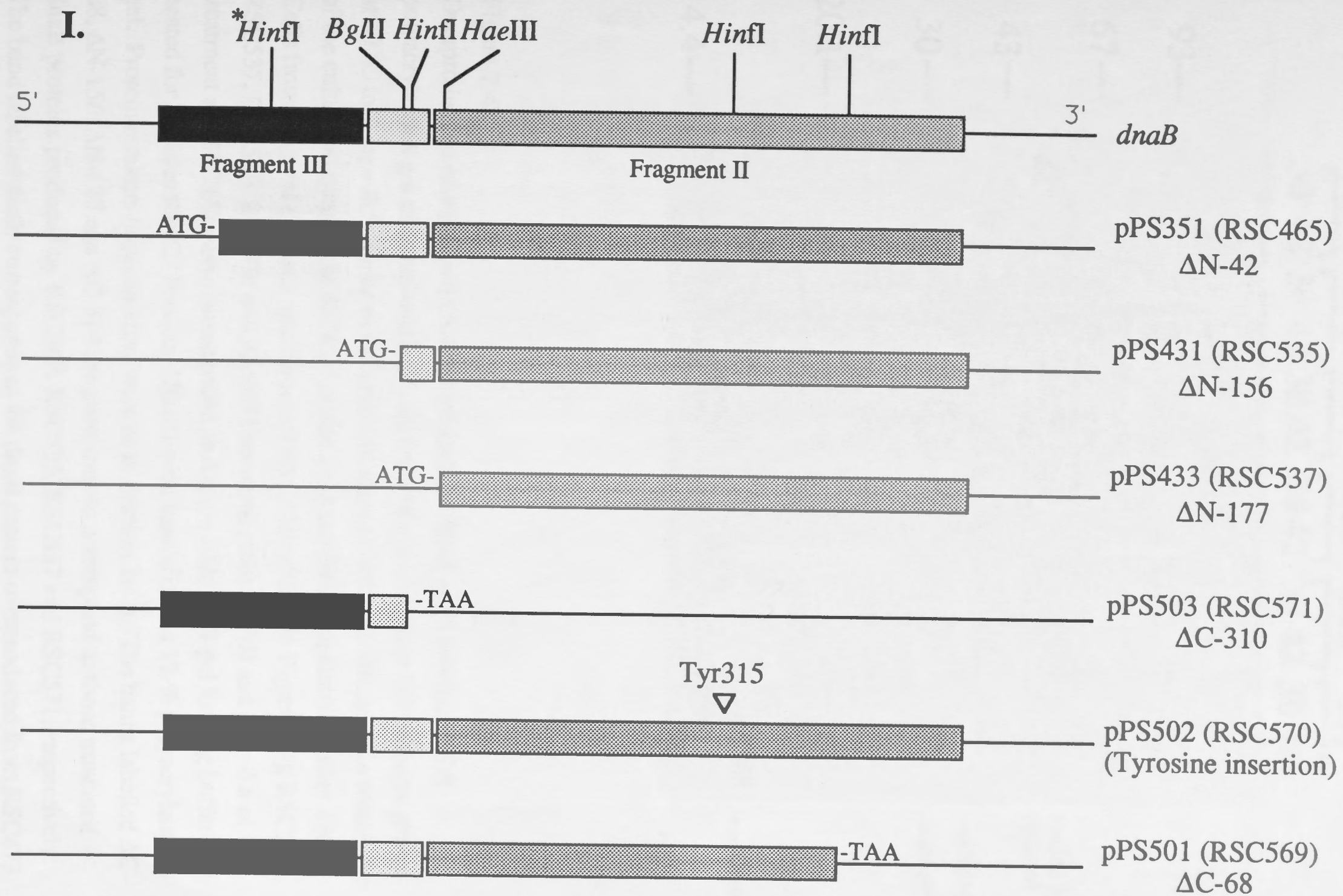


Figure 7.4A

Schematic representations of the coding regions of truncated dnaB proteins and their products generated by manipulation of restriction endonucleases. I. Shown is a representation of the coding region of the full length dnaB protein divided into regions predicted to encode the fragments isolated for partial tryptic hydrolysis of the protein (Nakayama *et al.*, 1984a). Restriction endonuclease sites listed were utilized as described (Sections 7.2.1 and 7.2.2). Products from these experiments included two plasmids directing NH₂-terminal deletions of dnaB (pPS431 and pPS433), two plasmids directing COOH-terminal deletions of dnaB (pPS503 and pPS501) and one plasmid which encodes a tyrosine insertion mutant (pPS501). The plasmid pPS351 was generated by a Bal31 deletion into the coding region of dnaB during experiments designed to engineer plasmids directing overproduction of the protein (Figure 4.7). The asterisk represents that *Hinf*I site for which no insertion mutant was recovered (Section 7.2.2). II. Displayed is a representation of the structured domains of the dnaB protein. Positions of mutations (I) leading to new translation initiation codons (**ATGXX) and stop codons (XXTAA**) are shown.



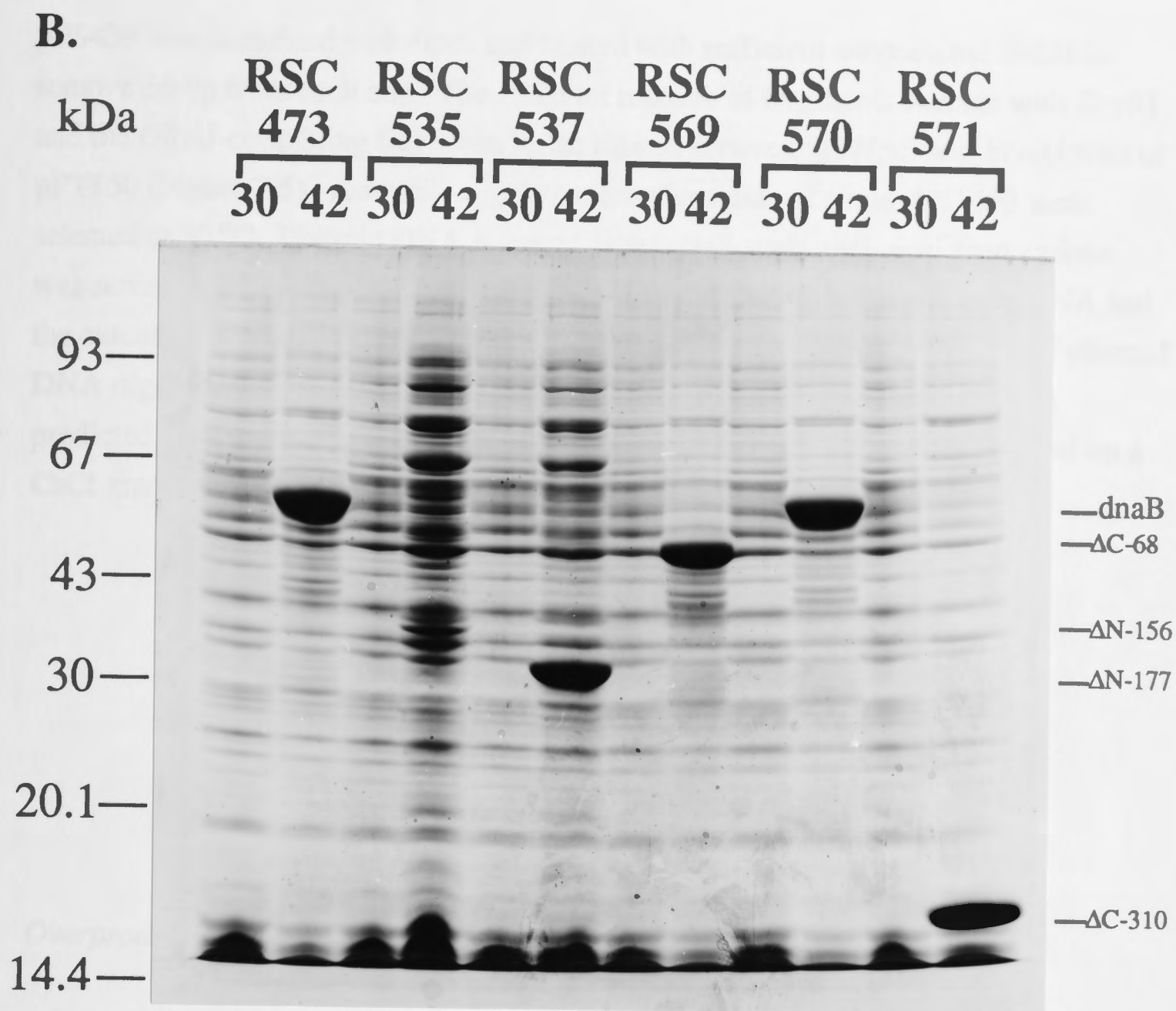


Figure 7.4B

Overproduction of NH₂- and COOH-terminally deleted dnaB proteins. SDS polyacrylamide gel electrophoresis of cell-free extracts. Cultures (20 ml) were grown at 30 °C to A₅₉₅ ~ 0.5. Cells were harvested from a portion (1 ml), and the remainder of the culture was shaken at 42 °C. Further 1-ml samples were removed after 4 h. Cells from the 1-ml portions of cultures of RSC473 (pPS359; Figure 4.7), RSC535, RSC537, RSC569, RSC570 and RSC571 removed prior to (30) and after 4 h of treatment at 42 °C (42) were resuspended to A₅₉₅ = 10 in SDS-gel loading buffer and heated for 2 min at 95 °C. Portions (20 µl) were loaded onto a 12-% polyacrylamide gel. Protein markers (sizes in kDa) were as in Section 2.4.3. The bands labelled ΔC-68, ΔN-156, ΔN-177 and ΔC-310 are presumed to correspond to those truncated dnaB proteins produced by RSC569, RSC535, RSC537 and RSC571, respectively. The band labelled dnaB corresponds to the dnaB protein overproduced from RSC473. The Tyr insertion mutant protein produced by RSC570 is also presumed to correspond closely to this band.

pPS439 was linearized with *HpaI* and treated with sufficient exonuclease Bal31 to remove 36 bp from each end. The resultant mixture of fragments was cut with *EcoRI* and the *ORF3*-containing fragments again ligated between the *HpaI* and *EcoRI* sites of pPT150 (Figure 7.5). Ampicillin resistant transformants of strain AN1459 were selected at 30 °C. Plasmid DNA prepared from small-scale analytical preparations was screened for the anticipated size (~5.7 kb) by *PstI* digestion of plasmid DNA and the extent of exonuclease digestion was estimated by 3' end-labelling studies of plasmid DNA digested with *EcoRV* and *BamHI* (see Section 2.3.11). Those plasmids predicted to contain a short RBS-ATG linker length (<16 bp) were then purified on a CsCl gradient and sequenced using the dideoxy method.

RESULTS

Overproduction of subunit domains of the dnaB protein

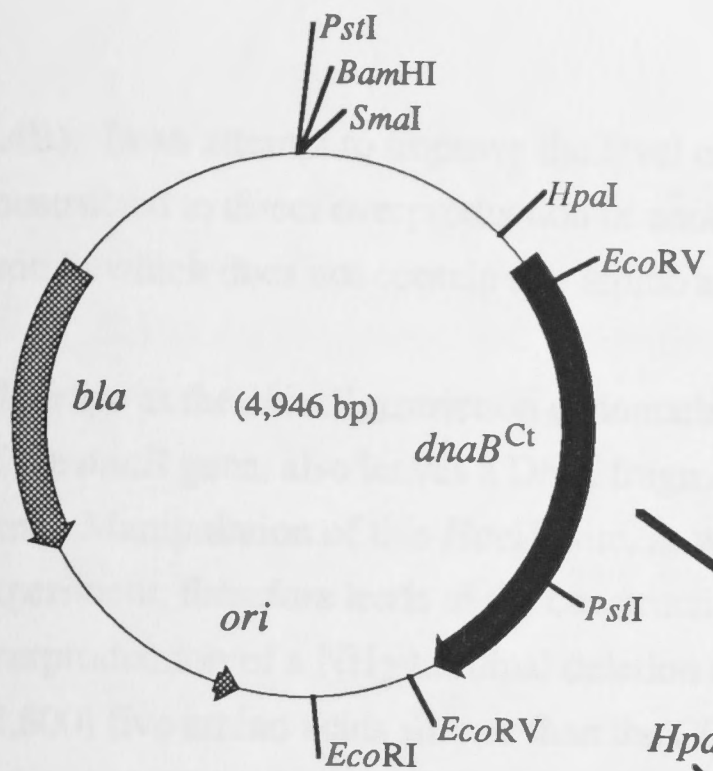
Cleavage at the unique *BglIII* restriction endonuclease site within the coding region of the *dnaB* gene, followed by repair of the ends with Klenow enzyme, produces a blunt end in-frame with the original reading frame of the *dnaB* gene. By fusing this end to an ATG translation initiation codon downstream of a synthetic ribosome-binding site, a new *dnaB* mutant was created encoding the COOH-terminus of dnaB protein commencing at Asp156 (Figures 7.1 and 7.4).

The vector used to generate the new ATG translation initiation codon was one of several RBS-ATG vectors (pND211-pND217; Figure 2.1) which, when linearized with *NcoI* and end-filled, results in a blunt end with the sequence 5'-.....CCATG^{-OH}. The terminal ATG can then be employed as a translation initiation codon for coding regions of genes inserted in frame.

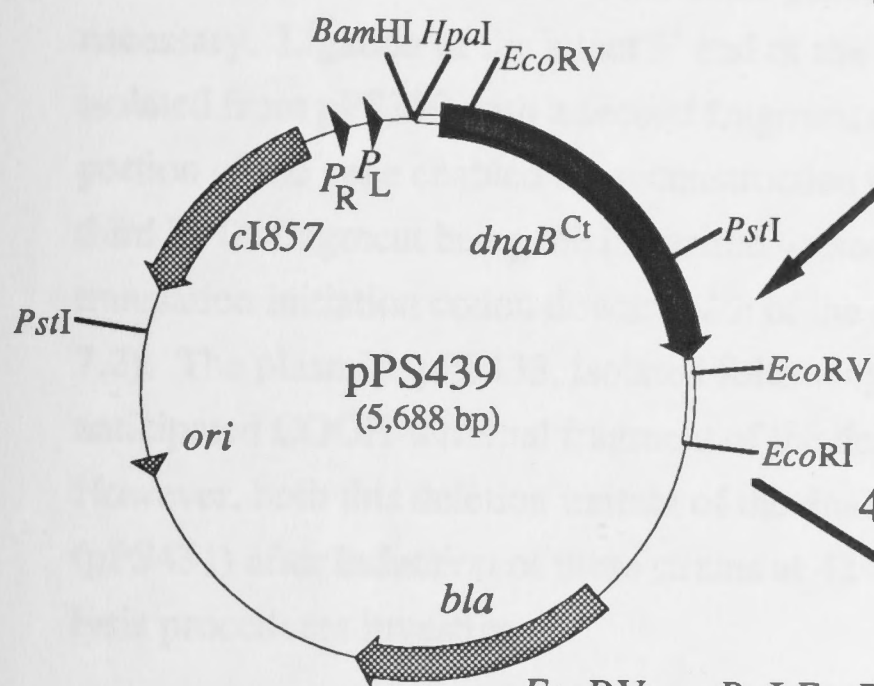
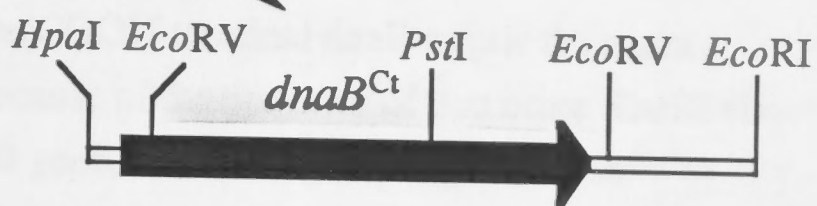
This strategy produced a strain (RSC 535) bearing the new plasmid construct that directed overproduction of an NH₂-terminal dnaB deletion (dnaB-Δ156, predicted *M_r* 35,400) following thermal induction at 42 °C. The deletion mutant, a protein commencing 16 amino acids shy of the NH₂-terminus of the proposed COOH-terminal domain of the dnaB protein (Figure 7.4A), was overproduced to significantly lower levels than produced from other overexpression plasmids (e.g. pPS359; Figure

Figure 7.5

Construction of a plasmid, pPS504, that is predicted to direct the overproduction of a Chlamydial *dnaB*-like protein. The *Pst*I sites used for screening of plasmids of anticipated size for the presence of inserts, and approximate end-points of *Bal*31 digestion are centered 917 bp from the 5' end of the *dnaB*^{ct} coding region and 1349 bp upstream of the *Bam*HI site in the vector. The *Eco*RV site used in 3'-end labelling experiments is located 40 bp downstream of the *dnaB*^{ct} ATG translation initiation codon. A second *Eco*RV site is located 1400 bp further downstream. The relevant sequences of plasmids selected from this experiment are given (Figure 7.6).

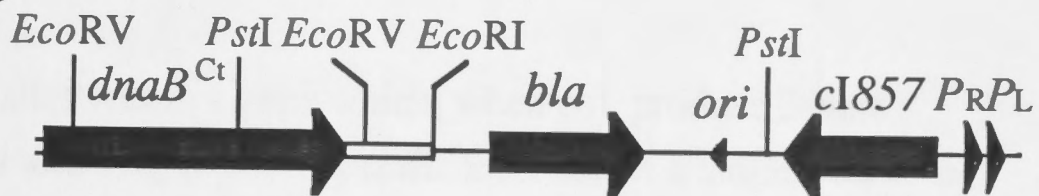


1. *Hpa*I/*Eco*RI
2. Fragment isolation

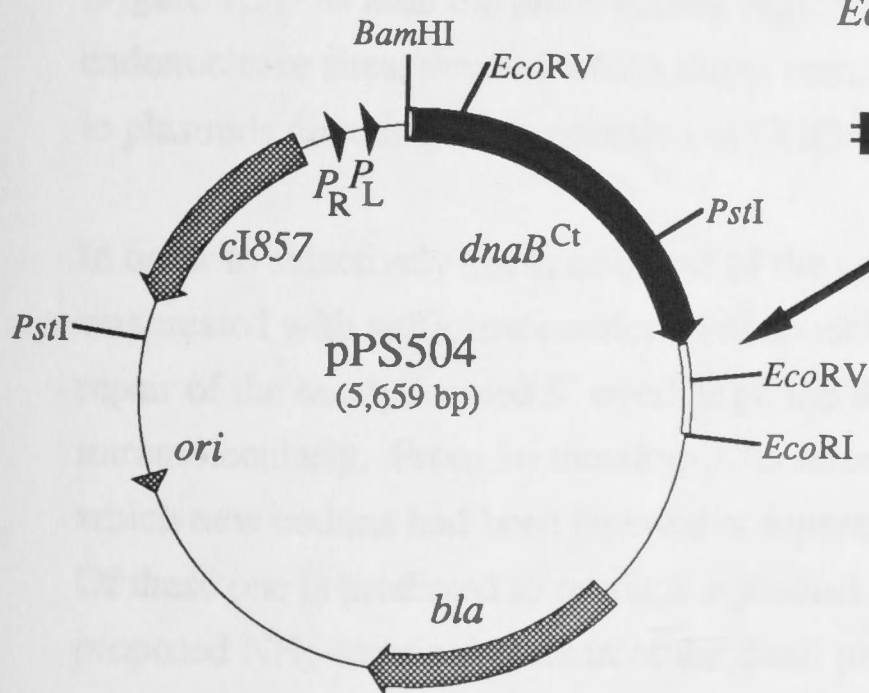
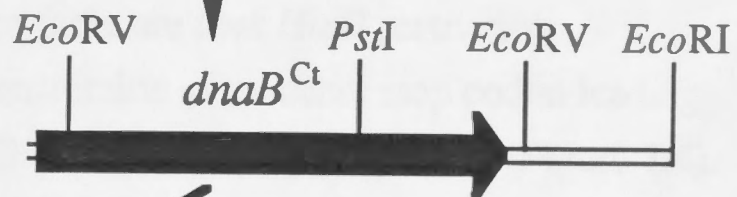


3. *Hpa*I/*Eco*RI-linearized pPT150 + ligase

4. *Hpa*I
5. *Bal*31



6. *Eco*RI
7. Fragment isolation



8. *Hpa*I/*Eco*RI-linearized pPT150 + ligase

7.4B). In an attempt to improve the level of overproduction, a second plasmid was constructed to direct overproduction of another NH₂-terminal deletion of the *dnaB* protein which does not contain any amino acids from the putative hinge region.

Cleavage at the *Hae*III restriction endonuclease site 532 ntds downstream of the 5' end of the *dnaB* gene, also leaves a DNA fragment whose blunt end lies in frame with the gene. Manipulation of this *Hae*III site, as with the *Bgl*III site in the previous experiment, therefore leads to the construction of a truncated *dnaB* gene directing the overproduction of a NH₂-terminal deletion of *dnaB* (*dnaB*- Δ 177, predicted M_r 32,800) five amino acids shorter than the COOH-terminal *dnaB* tryptic fragment isolated by Nakayama *et al.* (1984). Because of the presence of two other *Hae*III sites towards the 3'-OH terminus of the *dnaB* gene, a three fragment ligation was necessary. Ligation of the intact 3' end of the gene on an *Nco*I/*Eco*RI fragment isolated from pPS359 with a second fragment containing the *Hae*III/*Nco*I upstream portion of the gene enabled the reconstruction of the entire *dnaB* deletion mutant, the third DNA fragment being the linearized vector pND217 bearing the blunt-ended ATG translation initiation codon downstream of the synthetic ribosome-binding site (Figure 7.2). The plasmid, pPS433, isolated following this strategy, overproduced the anticipated COOH-terminal fragment of the *dnaB* protein to high-levels (Figure 7.4). However, both this deletion mutant of the *dnaB* protein and that produced by RSC535 (pPS431) after induction of these strains at 42 °C were totally insoluble following all lysis procedures investigated.

Restriction endonuclease sites within a gene which, when cut, produce 3-base overhangs on either strand allowing the non-specific insertion of a single codon in-frame following repair of ends and blunt-end religation (Figure 7.3). Using such a strategy with *Hinf*I has the added advantage in favourable sequences of enabling the easy insertion of an in-frame opal (TGA) or ochre (TAA) translation stop codon (Figure 7.3). Within the *dnaB* coding region there are four *Hinf*I restriction endonuclease sites, three of which direct construction of an ochre stop codon leading to plasmids directing the production of COOH-terminal deletion mutants (Figure 7.4).

In order to selectively cut at only one of the twelve *Hinf*I sites in pPS359, the plasmid was treated with sufficient restriction endonuclease just to linearize it. Following repair of the newly formed 5' overhangs, the linear plasmid was isolated and religated intramolecularly. From 36 transformants screened, three plasmids were isolated in which new codons had been inserted at separate *Hinf*I sites within *dnaB* (Figure 7.4). Of these one is predicted to produce a product (M_r = 19,300) that comprises the proposed NH₂-terminal domain of the *dnaB* protein. A second larger NH₂-terminal

domain ($M_r = 46,200$) is predicted following repair of the third *HinfI* site in the *dnaB* gene. Cleavage and end-filling of the fourth *HinfI* site is predicted to result in the insertion of a tyrosine codon (Tyr315*) between those that encode residues Asp314 and Ser315 of the *dnaB* protein. The fourth possible plasmid construct, for which the insertion of an ochre stop codon following that of Glu89 was predicted, was not isolated.

Each isolated plasmid directed the overproduction of its respective mutant product in cells following thermal induction at 42 °C (Figure 7.4), although only protein, produced from strain RSC571 (containing pPS503), was soluble following trial lysis experiments. This strain is the one predicted to direct overproduction of a protein that includes the NH₂-terminal domain isolated following partial tryptic hydrolysis (Nakayama *et al.*, 1984a), and most of the proposed hinge region between the domains (Figure 7.4).

Overproduction of dnaB from C. trachomatis

A plasmid designed to overproduce the *dnaB* homologue from *Chlamydia trachomatis* was constructed in an almost identical procedure to that used for the construction of pPS359, the *E. coli dnaB* overexpression plasmid (Section 4.2.3). Following treatment of an appropriate DNA fragment, containing the entire coding region of the *C. trachomatis* gene (*dnaB^{ct}*), with exonuclease Bal31 to remove up to 36 bp of DNA that precedes the ATG translation initiation codon, the mixed fragment pool was digested with *EcoRI*, thus ensuring the insertion of the gene in the correct orientation in the vector pPT150. Forty-eight ampicillin resistant transformants were selected for further screening by restriction endonuclease digests of their plasmid DNA, of which three were isolated for dideoxy sequencing (Figure 7.5). Two of these had RBS-ATG linker lengths that were predicted to allow overproduction of the full length protein (predicted $M_r = 51,300$; Figure 7.6). A third, with the synthetic ribosome-binding site of the vector fused sufficiently close to the ATG codon normally encoding Met11 that it ought to become a new translation initiation codon, was predicted to overproduce an NH₂-terminal deletion (*dnaB^{ct}*-Δ11, predicted $M_r = 50,100$). Neither this nor the other two plasmid constructs successfully overproduced a protein of the anticipated size following thermal induction of strains containing these plasmids (results not shown).

These results prompted concerns that differences in codon preference between *C. trachomatis* and *E. coli* (Wada *et al.*, 1990) may also represent the relative amounts of

Strain (plasmid)	Nucleotide sequence			
pCTT1	...GCTCGG-TATTTTTTTT-ATATAAAC	ATG	AAA	AAC...
			met	lys asn
		<i>Bam</i> HI	Ribosome- Binding Site	Linker
RSC504 (pPS572)	...GGATCC-TAAGGAGGT-TTATAAAC	ATG	AAA	ACT...
		¹ met	lys	thr
RSC505 (pPS573)	...GGATCC-TAAGGAGGT-TAC-----	ATG	AAA	ACT...
		¹ met	lys	thr
RSC506 (pPS574)	...GGATCC-TAAGGAGGT-TGC-----	ATG	CAA	GAT...
		met ¹²	gln	asp

Figure 7.6

Nucleotide sequences determined for three plasmids in the region of the *dnaB*^{ct} translation initiation codon. The sequence above is that of the *dnaB* gene from pCTT1. Neither plasmid containing only a short linker between the synthetic ribosome-binding site and the vector of the ATG translation initiation codon of *dnaB*^{ct} directed detectable expression of a protein corresponding to the expected size of *dnaB*^{ct} (51.3 kDa, not shown). A third plasmid predicted to direct transcription of a truncated *dnaB*^{ct} gene was also incapable of expressing the corresponding protein to a detectable level (not shown).

their isoaccepting tRNAs *in vivo* and that gene expression might thus be severely compromised. In an attempt to establish whether the failure of these strains to overproduce the *dnaB*^{ct} protein and its NH₂-terminal deletion was limited by their capacity to provide cognate tRNAs, the plasmid pPS504 was also expressed in two strains that simultaneously overproduced the tRNA^{Arg}_{AGA} (the *dnaY* (*argU*) gene product; Figure 3.6). However, neither strain produced quantities of the *dnaB*^{ct} protein detectable by SDS-PAGE (results not shown).

DISCUSSION

The overproduction of a series of COOH- and NH₂-terminal deletion mutants of *dnaB* was undertaken with a view to isolating the individual activities of the protein. The *dnaB* protein is believed to have two structured domains separated by a hinge region. By appropriate use of restriction endonuclease sites within the gene, new translation stop codons and ATG translation initiation sites were inserted to create genes encoding polypeptides that mimic the NH₂- and COOH-terminal domains of the protein (Figure 7.4).

Of the two COOH-terminal polypeptides (Figure 7.4) the larger one that contains a number of amino acids from the proposed hinge region of the *dnaB* protein was overproduced to only a low level. The sequences responsible for the regulation of expression of both these polypeptides are identical. It can only be presumed therefore, that the additional sixty-nine nucleotides of the reading frame encoding the longer polypeptide affects the efficiency of translation of the message by the ribosome. The regions of the mRNA encoding the hinge and surrounding sequences are predicted to contain a considerable number of small inverted repeats which may, in part, contribute to the low translational efficiency by forming secondary structures. The production of the NH₂-terminal polypeptide, however, does not appear to be limited by these sequences which encode its COOH-terminal end. In the case of the *dnaB* protein, this could suggest that the slow rate of translation in this region may enable complete folding of the NH₂-terminal domain prior to translation of the COOH-terminal domain.

The engineered NH₂-terminal domain was quite soluble, although the COOH-terminal

polypeptides were not. This difference in solubility may reflect the relative insolubility of the overproduced native protein (Chapter 4), where the COOH-terminal domain may be responsible for formation of insoluble proteinaceous aggregates following thermal induction. Co-overproduction of the native dnaB protein with dnaC rendered the dnaB protein soluble, therefore implicating the COOH-terminal domain of dnaB in interaction with the dnaC protein. The dnaC protein has previously been presumed to interact with the NH₂-terminal domain of dnaB (Nakayama *et al.*, 1984a), although the present results and the recent analysis of a novel temperature-sensitive dnaB mutation (Chang *et al.*, 1991) now suggest otherwise. Experiments designed to overproduce the COOH-terminal domain in the presence of dnaC protein using strategies similar to those outlined in Chapter 4 are currently underway. Proposed future research also includes the purification of the fragments of the dnaB protein corresponding to identified domains, the study of their functions in appropriate assays, structural studies on them and experiments designed to reconstitute the native protein from discrete fragments.

Overproduction of dnaB protein from C. trachomatis

Chlamydia trachomatis is an obligate intracellular parasitic bacterium that causes a wide range of infections in humans. It possesses all the cell machinery for prokaryotic DNA, RNA and protein synthesis, but unlike other gram-negative bacteria (*e.g. E. coli*), it lacks the ability to generate high-energy compounds for metabolism. Both the LGV biovar, which causes lymphogranuloma venereum, and the trachoma biovar, which causes ocular and genital infections, possess a small homologous plasmid (Sriprakash and Macavoy, 1987; Hatt *et al.*, 1988). As the chlamydial genome is small (Kingsbury, 1969) and as the plasmid is ubiquitous, it is likely that the plasmid plays an essential role in the biology of the parasite. Analysis of the open reading frames of both plasmids reveal a gene encoding a protein whose COOH-terminus is highly homologous to that of the dnaB protein of *E. coli* (Appendix B), and itself may have functions in *Chlamydia* analogous to those of the dnaB protein.

The Chlamydial *dnaB* gene from pCTT1, the plasmid isolated from *C. trachomatis*, trachoma biovar (Sriprakash and Macavoy, 1987), was manipulated in ways similar to methods used in the construction of the dnaB overexpression vector, pPS359. The natural ribosome-binding site of the gene was removed by treatment of an appropriate DNA fragment with the double-stranded exonuclease Bal31. The gene was then provided with a new ribosome-binding site perfectly complementary to the 3'-OH terminus of *E. coli* 16-S rRNA and transcription was placed under the direction of the

tandem strong bacteriophage λ promoters P_R and P_L in pPT150, a derivative of the high-copy number plasmid pUC9 (Elvin *et al.*, 1990). Screening by DNA sequence determination yielded a number of plasmids that ought to have directed high-level expression of the dnaB^{ct} protein (Figure 7.6). However, the overproduced protein was not detected by SDS-PAGE.

Concern as to the failure of these plasmids to direct the overproduction of the dnaB^{ct} protein were focused on differences noted in the sequence of the genes encoding dnaB^{ct} from pCTT1 (Sriprakash and Macavoy, 1987) and pLGV440 (Hatt *et al.*, 1988). A single nucleotide alteration in each gene affected the size of the NH_2 -terminus of the LGV dnaB^{ct} protein and the COOH -terminus of the trachoma dnaB^{ct} protein (Appendix B; footnote). However, the translation initiation codon assumed for experiments designed to improve the production of the dnaB^{ct} protein was confirmed (Dr K.A. Sriprakash, personal communication). Mapping of transcripts is being carried out to confirm that protein synthesis is not limited by mRNA production or stability, although a more likely explanation is that translation is limited by the capacity of *E. coli* to provide cognate tRNA's for codons common on *C. trachomatis* genes but rare in *E. coli*.

Nucleotide sequences, which are becoming available in increasing numbers, have clearly indicated that the spectrum of synonymic codon preference appears to differ from organism to organism (Wada *et al.*, 1990; Andersson and Kurland, 1990). It has already been determined that, within an organism, certain synonymous codons are preferred to others. This has led to the completely unconfirmed suggestion that genes containing these codons (*e.g.*, the *dnaG* gene; Chapter 3) may be regulated at the level of translation through the availability of their isoaccepting tRNA's (Ernst, 1988; Brinkmann *et al.*, 1989). *C. trachomatis* utilizes several codons which occur only rarely in the *E. coli* genome, and this quite extreme bias is maintained within the coding region of the dnaB^{ct} gene (Table 7.1). The overproduction of the scarce *E. coli* tRNA^{Arg}_{AGA} did not appear to improve the expression of the dnaB^{ct} gene suggesting that the translational efficiency may still be limited by other rare *E. coli* tRNAs (notably those recognizing the CUA_(Leu) and AUA_(Ile) codons). Organism-specific preference for tRNAs may therefore make the expression of Chlamydial proteins difficult in *E. coli* hosts. However, as evident for Table 7.1, the codon usage of the yeast *Saccharomyces cerevisiae* parallels that of *Chlamydia spp.* rather nicely. Since good methods are now available for gene expression in yeast, there is some prospect that we may soon be able to express dnaB^{ct} and other replication genes from *Chlamydia* (Dr K.A. Sriprakash, personal communication) in that organism.

Table 7.1

Codon usage in sequenced *E. coli* and *C. trachomatis* genes and the frequency of occurrence within the *dnaB* coding regions from these organisms. The frequency (per 1000 residues) of codon use in each organism was taken from Wada *et al.* (1990). Codon usage in *dnaB* from *E. coli* (471 amino acids) and *dnaB^{ct}* of pCTT1 from *C. trachomatis* (451 amino acids) was calculated on the same basis. Those codons for which the supply of cognate tRNAs may be insufficient for high-level expression of *dnaB^{ct}* in *E. coli* are indicated by shading. It is noteworthy the *S. cerevisiae* displays a similar codon bias to *C. trachomatis* and may perhaps be a more appropriate host for expression of Chlamydial genes such as *dnaB^{ct}*.

Codon	<i>E. coli</i>	<i>C. trachomatis</i>	<i>S. cerevisiae</i>	<i>dnaB</i> gene	<i>dnaB^{Ct}</i> gene	Codon	<i>E. coli</i>	<i>C. trachomatis</i>	<i>S. cerevisiae</i>	<i>dnaB</i> gene	<i>dnaB^{Ct}</i> gene
	Occurence / 1000 residues						Occurence / 1000 residues				
Arg						Gly					
CGA	3.0	6.4	2.3	0.0	8.9	GGA	6.6	25.9	8.9	2.1	19.9
CGC	21.7	7.0	2.0	33.9	6.6	GGC	30.5	7.0	8.9	21.2	4.4
CGG	4.4	2.9	1.1	0.0	8.9	GGG	9.4	10.4	5.1	19.1	8.9
CGU	25.5	9.0	7.5	44.5	17.7	GGU	28.6	14.2	34.9	21.2	15.5
AGA	2.0	17.3	24.0	0.0	28.7	Val					
AGG	1.2	5.7	7.5	0.0	8.9	GUA	12.2	16.9	10.0	12.7	13.3
Leu						GUC	14.0	7.8	14.9	12.7	0.0
CUA	2.9	12.7	11.8	2.1	11.1	GUG	24.7	9.7	9.5	23.3	11.1
CUC	9.6	6.4	4.1	12.7	6.6	GUU	20.8	22.1	26.6	10.6	31.0
CUG	54.8	5.9	8.3	46.6	8.9	Lys					
CUU	9.9	17.1	9.6	8.5	17.9	AAA	37.4	53.4	37.7	36.0	48.6
UUA	10.2	31.7	24.5	4.2	33.1	AAG	12.1	19.9	35.2	2.1	15.5
UUG	11.2	23.5	32.1	19.1	26.5	Asn					
Ser						AAC	24.6	18.3	25.8	36.0	6.6
UCA	6.4	12.0	15.6	4.2	15.5	AAU	15.9	31.1	31.4	8.5	48.6
UCC	9.6	13.4	14.4	17.0	11.1	Gln					
UCG	7.8	6.6	6.5	23.3	15.5	CAA	12.9	28.0	29.8	12.7	19.9
UCU	10.5	34.7	24.6	10.6	35.4	CAG	30.2	11.2	10.4	36.0	13.3
AGC	14.9	9.4	7.1	8.5	13.3	His					
AGU	7.2	11.3	11.7	8.5	8.9	CAC	10.8	4.6	8.2	10.6	0.0
Thr						CAU	11.3	8.6	12.3	2.1	15.5
ACA	6.4	22.0	15.6	4.2	8.9	ACG	12.2	5.6	6.7	10.6	2.2
ACC	24.2	8.3	13.9	19.1	2.2	Glu					
ACU	10.6	24.7	22.0	6.4	13.3	GAA	44.3	41.2	48.9	59.3	44.2
Pro						GAG	19.5	17.2	16.9	19.1	26.5
CCA	8.1	9.3	21.4	14.8	8.9	Asp					
CCC	4.0	3.6	5.9	4.2	2.2	GAC	22.5	17.3	22.3	33.9	19.9
CCG	24.1	2.7	4.1	21.2	2.2	GAU	31.8	34.3	37.0	44.5	44.3
CCU	6.4	16.4	12.8	4.2	4.4	Tyr					
Ala						UAC	13.5	13.2	16.5	6.4	8.9
GCA	20.4	22.7	15.3	14.8	15.5	UAU	15.0	15.3	16.6	14.8	15.5
GCC	23.0	8.9	15.5	21.2	2.2	Phe					
GCG	32.9	8.7	5.1	42.3	11.1	UUC	18.3	19.1	20.0	6.4	31.0
GCU	17.8	38.7	28.3	12.7	33.1	UUU	18.6	28.4	23.2	19.1	22.1
Cys						Ile					
UGC	6.1	5.5	3.7	0.0	2.2	AUA	3.8	19.1	12.8	0.0	44.2
UGU	4.7	7.1	7.6	0.0	8.9	AUC	27.6	20.3	18.4	50.8	19.9
Met						AUU	27.0	33.5	31.1	21.2	28.7
AUG	26.6	20.9	21.5	29.7	15.5	Ter					
Trp						UAA	2.0	1.6	1.1	2.1	2.2
UGG	12.7	10.5	10.5	6.4	4.4	UAG	0.2	0.5	0.4	0.0	0.0
						UGA	0.8	1.0	0.6	0.0	0.0

CONCLUDING STATEMENT

A remarkable device for RNA primer synthesis during chromosomal replication in *Escherichia coli* is the primosome, an assembly of seven distinctive proteins including primase, the *dnaB*, *dnaC* and *dnaT* gene products and the PriA, PriB and PriC proteins. This complex is believed responsible at least for repeated initiations on the lagging strand in discontinuous DNA synthesis at replication forks. The primosome, therefore, presents a particularly useful model system for the study of the mechanisms of protein-protein and protein-nucleic acid interactions.

The specific aim of the work presented in this thesis was the successful overexpression of primosomal genes and purification of sufficient quantities of the proteins for chemical and structural studies. High-level expression of primase and the *dnaB*, *dnaC* and *dnaT* proteins was achieved by using recombinant DNA technology to replace the DNA sequences responsible for the normal control of expression of their genes. In each case, the natural promoter region and ribosome-binding site of the gene was removed by exonuclease digestion. Alternative strategies were then employed to provide a new ribosome-binding site perfectly complementary to the 3'-OH terminus of 16-S ribosomal RNA just upstream of the translation initiation codon of the gene, and to place transcription of the gene under the control of the tandem strong bacteriophage λ promoters P_R and P_L .

Overproduction and large-scale purification of three of the primosomal proteins has made the study of the structure and function of these proteins feasible. As an example of the importance of obtaining large stocks of readily available proteins, a highly purified preparation of primase from one such source enabled, for the first time, the identification and determination of stoichiometry of bound zinc. Future investigations will centre on metal ion replacement studies and studies directed at identifying donor ligands involved in metal ion co-ordination, as well as the crystallization and structural determination of this enzyme. This should present a perfect opportunity to investigate the role of zinc in a much broader spectrum of related proteins, since no crystal structure of a "zinc finger" protein has yet been reported.

The similar overproduction of *dnaB* and *dnaC* and isolation of large quantities of these proteins will also lead to studies of their structure and chemistry of action. This has already been facilitated in a very limited way by site directed mutagenesis of a specific motif in the *dnaB* protein, and an initial study of the interaction of the *dnaB* protein with *dnaC*. The ongoing study of the structural domains of the *dnaB* protein will add

invaluable data to this cause.

Of greater importance than just the physical data that will inevitably be obtained for these proteins is that the primosomal proteins were specifically selected for the unique opportunity of studying an integrated protein complex which, through the coordination of the many functions of its individual components, serves a specific function in the biology and chemistry of DNA. This has been the driving force behind the bias in this thesis toward investigations of the dnaB protein. This multifunctional protein marshalls the primosome and its functions and interacts functionally and probably also physically with most, if not all, of the other six primosomal proteins.

The isolation of large quantities of an intact dnaB-dnaC complex will enable initiation of a thorough investigation of the roles and interactions performed by each of the other primosomal proteins. With this and the recent isolation of genes encoding the PriA, PriB and PriC proteins, the prospect for isolation of all the known primosomal proteins in quantities sufficient for structural studies in reconstituted sub-assemblies is therefore now rather good.

The range of topics covered in this work has been of necessity divided, since it represents the very first stage of an ongoing long-term program. A solid foundation has now been set in place, and real progress in understanding the structural basis for the mechanics of operation of this remarkable assembly of proteins should soon follow.

490 510 530
AspLeuLeuAspLeuArgGluSerArgValPheLysIleAlaGluSerArgAlaAsnLys
GATCTGCTGGATCTGGCTGAATCCCGCGTCTTTAAAATTGCCGAAAGTCGTGCGAACAAA
A G G G CC G C G AC G C
Glu Gln Asn

550 570 590
AspGluGlyProLysAsnIleAlaAspValLeuAspAlaThrValAlaArgIleGluGln
GACGAAGGGCCGAAGAACATCGCCGATGTGCTCGACGCAACCGTGGCGCGTATTGAGCAG
T A G A C GA T C
Ser AspGlnIle

610 630 650
LeuPheGlnGlnProHisAspGlyValThrGlyValAsnThrGlyTyrAspAspLeuAsn
TTGTTTCAGCAGCCACACGATGGCGTTACCGGGGTAAACACCGGTTATGACGATCTCAAC
C A G A C GG T C C G T
Asp Gln

670 690 710
LysLysThrAlaGlyLeuGlnProSerAspLeuIleIleValAlaAlaArgProSerMet
AAAAAAACCGCTGGCTTGCAGCCGTCGGATTTGATCATCGTCGCCGCGCGTCCGTCGATG
G G A G A GT T G C C
Arg

730 750 770
GlyLysThrThrPheAlaMetAsnLeuValGluAsnAlaAlaMetLeuGlnAspLysPro
GGTAAAACAACATTTGCGATGAACCTCGTCGAAAACGCGGCGATGTTGCAGGATAAACCG
C T TG T A G
Cys

790 810 830
ValLeuIlePheSerLeuGluMetProSerGluGlnIleMetMetArgSerLeuAlaSer
GTACTTATCTTCTCGCTGGAGATGCCATCAGAACAGATCATGATGCGTTCTCTGGCGTCG
G TAGT CGGC ATG C
Gly Met

850 870 890
LeuSerArgValAspGlnThrLysIleArgThrGlyGlnLeuAspAspGluAspTrpAla
CTGTCGCGCGTTGACCAGACTAAAATCCGTACCGGGCAGCTCGATGACGAAGACTGGGCG
C C T ACGT T T A T T
Arg

910 930 950
ArgIleSerGlyThrMetGlyIleLeuLeuGluLysArgAsnIleTyrIleAspAspSer
CGCATTTCCGGCACCATGGGTATTTTGCTCGAAAAACGCAATATCTATATCGATGACTCC
A C G C C G G G T C T C
Met

970 990 1010
SerGlyLeuThrProThrGluValArgSerArgAlaArgArgIleAlaArgGluHisGly
TCCGGCCTGACGCCAACGGAAGTGCGTTCCCGCGCACGCCGTATTGCCCGTGAACACGGC
A T T G A C G G G T T C C
Phe

1030 1050 1070
GlyIleGlyLeuIleMetIleAspTyrLeuGlnLeuMetArgValProAlaLeuSerAsp
GGCATCGGGCTTATCATGATCGACTACCTGCAACTGATGCGCGTACCGGCGCTTTCCGAT
GT AA T G T T G T T
LeuSer Ser

1090 1110 1130
AsnArgThrLeuGluIleAlaGluIleSerArgSerLeuLysAlaLeuAlaLysGluLeu
AACCGTACGCTGGAAATTGCAGAAATCTCTCGCTCGCTGAAAGCACTGGCGAAAGAACTG
T C C T T G G C

1150 1170 1190
AsnValProValValAlaLeuSerGlnLeuAsnArgSerLeuGluGlnArgAlaAspLys
AACGTGCCGGTGGTGGCGCTGTCCCAGTTGAACCGTTCTCTGGAACAACGTGCCGACAAA
C G C A G C T C C C G T
Gln

1210 1230 1250
ArgProValAsnSerAspLeuArgGluSerGlySerIleGluGlnAspAlaAspLeuIle
CGCCCGGTCAACTCCGACCTGCGTGAATCTGGCTCTATCGAGCAGGATGCGGACTTGATC
T G C T A C A T

1270 1290 1310
MetPheIleTyrArgAspGluValTyrHisGluAsnSerAspLeuLysGlyIleAlaGl
u
ATGTTTATCTATCGTGATGAGGTGTATCACGAAAACAGTGATTTAAAAGGCATCGCGGAA
C T G C C T T

1330 1350 1370
IleIleIleGlyLysGlnArgAsnGlyProIleGlyThrValArgLeuThrPheAsnGly
ATTATTATCGGTAAACAACGTAAACGGCCCAATCGGGACGGTACGCCTGACCTTTAACGGT
T G T T T T G C T

1390 1410 1430
GlnTrpSerArgPheAspAsnTyrAlaGlyProGlnTyrAspAspGlu*****
CAGTGGTCGCGCTTCGACAACCTATGCGGGGCCGACGTACGACGACGAATAATAATTATTT
A T A A T T G CTC CCG C

1450 1470 *alr* →
TATGAATTAGGTAATTAAAGCAAACACTTATCAAGGAACAAATGCAAGCGGCAACTGTT
AT -----CTT A TT A C
MetGlnAlaAlaThrVal

1510 1530 1550
ValIleAsnArgArgAlaLeuArgHisAsnLeuGlnArgLeuArgGluLeuAlaProAla
GTGATTAACCGCCGCGCTCTGCGACACAACCTGCAACGTCTTCGTGAACTGGCGCCTGCC

1570 1590 1610
SerLysMetValAlaValValLysAlaAsnAlaTyrGlyHisGlyLeuLeuGluThrAla
AGTAAAATGGTTGCGGTGGTGAAGCGAACGCTTATGGTCACGGTCTTCTTGAGACCGCG
GC A G
Leu

1630 1650 1670
ArgThrLeuProAspAlaAspAlaPheGlyValAlaArgLeuGluGluAlaLeuArgLeu
CGAACGCTCCCCGATGCTGACGCCTTTGGCGTAGCCCGTCTCGAAGAAGCTCTGCGACTG
T T G G T G A T

1690 1710 1730
 ArgAlaGlyGlyIleThrLysProValLeuLeuGluGlyPhePheAspAlaArgAsp
 CGTGC GGGGGAATCACCAAACCTGTACTGTTACTCGAAGGCTTTTTTGTATGCCAGAGAT
 A C G GC G AA C C G G G T C C GCC
 Gln Ile Ala

1750 1770 1790
 LeuProThrIleSerAlaGlnHisPheHisThrAlaValHisAsnGluGluGlnLeuAla
 CTGCCGACGATTTCTGCCCAACATTTTCATACGGCCGTGCATAACGAAGAACAGCTGGCT
 T C C G TG C G C A TC G T C
 CysLeu Gln

1810 1830 1850
 AlaLeuGluGluAlaSerLeuAspGluProValThrValTrpMetLysLeuAspThrGly
 GCGCTGGAAGAGGCTAGCCTGGACGAGCCGGTTACCGTCTGGATGAAACTCGATACCGGT
 C G C TGGAG CG A G G G
 AlaValGlu Ala

1870 1890 1910
 MetHisArgLeuGlyValArgProGluGlnAlaGluAlaPheTyrHisArgLeuThrGln
 ATGCACCGTCTGGGCGTAAGGCCGGAACAGGCTGAGGCGTTTTATCATCGCCTGACCCAG
 T C GC T C G G C C G T G C
 Glu Gln His

1930 1950 1970
 CysLysAsnValArgGlnProValAsnIleValSerHisPheAlaArgAlaAspGluPro
 TGCAAAAACGTTTCGTCAGCCGGTGAATATCGTCAGCCATTTTTCGCGCGCGGATGAACCA
 T T A C C T G G

1990 2010 2030
 LysCysGlyAlaThrGluLysGlnLeuAlaIlePheAsnThrPheCysGluGlyLysPro
 AAATGTGGCGCAACCGAGAAACAACCTCGCTATCTTTAATACCTTTTTCGGAAGGCAAACCT
 G C T AC T G AC T G C TC G T C
 Glu His Asp Ala Gln

2050 2070 2090
 GlyGlnArgSerIleAlaAlaSerGlyGlyIleLeuLeuTrpProGlnSerHisPheAsp
 GGTCAACGTTCCATTGCCGCGTCGGGTGGCATTCTGCTGTGGCCACAGTCGCATTTTGAC
 GC C T T C T C G T C

2110 2130 2150
 TrpValArgProGlyIleXX
 TGGGTGCGCCCGGGCATCXX
 C T ATTTTGTATGGCGTATCGCCGCTGGAGCACAAACCTGGGGG
 Ala IleLeuTyrGlyValSerProLeuGluHisLysProTrpGly

2170 2190 2210
 XXX
 XXX
 CCGGATTTTGGTTTTTCAGCCGGTGATGTCCTTAACCTCCAGTTTGATCGCGGTGCGTGAC
 ProAspPheGlyPheGlnProValMetSerLeuThrSerSerLeuIleAlaValArgAsp

2230 2250 2270
 XXX
 XXX
 CACAAAGCGGGCGAACC GGCTACGGCGGGACATGGGTGAGTGAGCGCGACACGCGC

Physical map of the proposed *dnaB-alr* operon at 92 minutes on the *E. coli* chromosome.

Nucleotide numbering starts at the *dnaB* translation initiating codon. The complete length of the *alr* gene (1134 bp) in *E. coli* was deduced from the published sequence of *E. coli tyrB* (Kuramitsu *et al.*, 1985), the homologous *S. typhimurium alr* gene (Galakatos *et al.*, 1986) and the sequence of pPS359. The *dnaB* gene is that of Nakayama *et al.* (1984a). An open reading frame 83 bp upstream of the *dnaB* gene in *E. coli* encoding a 38-kDa protein has been identified (N.E. Dixon, personal communication) and its direction of transcription suggests that the *dnaB* protein at least is transcribed from its own promoter, of which there are a number of candidate sequences (not identified). A GC-rich palindrome (opposing arrows above the nucleotide sequence) followed by a long stretch of consecutive AT bp downstream of the *alr* gene is characteristic of *rho*-independent transcription termination. This potential transcription termination site may provide termination for transcripts initiated by the *dnaB* promoter sequences.

The three letter code amino acid sequence for the corresponding nucleotide sequence is shown. The nucleotide and three letter amino acid sequence above is that of *E. coli*. Differences between the sequence from *E. coli* and that of *S. typhimurium* are denoted below the *E. coli* sequence. The start sites and direction of transcription of genes are indicated by thick arrows. Nucleotide sequences that have not been determined (X) or have been deleted (-) are indicated.

SEQ F:

G K P M T Y N V W N F G

Fragment III

25

SEQ A:	M A G N K P F N K Q Q A E P - - R	E R D P Q V A G L K V P P
SEQ B:	M A G N K P F N K P Q T D A - - R	D R D P Q V A G I K V P P
SEQ C:	- - - - -	- - - - - M K - - -
SEQ D:	- - - - -	- - - - - M K - - -
SEQ E:	- - - - -	- - - - - M R - - -

SEQ F: E S N G R - Y S A L T A R G I S K E T C - Q K A G V V Y W I

50

SEQ A:	- H S - I E A E - Q - - - - S V L G G L M L D N E R W D D V
SEQ B:	- H S - I E A E - Q - - - - S V L G G L M L D N E R W D D V
SEQ C:	T R S E I E N R M Q D I E Y A L L G K A L I F E D S T E Y I
SEQ D:	T R S E I E N R M Q D I E Y A L L G K A L I F E D S T E Y I
SEQ E:	- - - - - Q D I E - A S V I G G L L I G G L T P T A

SEQ F: A K V D G V M Y - Q V A D Y R D Q N G N I V S Q - - - - -

75

SEQ A:	A E R V V A D D F Y T R P H R H I F T E M A R L Q E S G S P
SEQ B:	A E R V V A E D F Y T R P H R H I F T E M G R L Q E S G S P
SEQ C:	L R Q L A N Y E F K C S H H K N I F I V F K Y L K D N G L P
SEQ D:	L R Q L A N Y E F K C S H H K N I F I V F K Y L K D N G L P
SEQ E:	S D V L A T L E P E A F S I P L Y R K A F E V I R K Q - - A

SEQ F: - - - - - K V R D K D K N F K T T G S H K S D A L F

SEQ A:	M D L I T - L A E S L - - - - - E R Q G - - Q L - - D -
SEQ B:	I D L I T - L A E S L - - - - - E R Q G - - Q L - - D -
SEQ C:	I T K - S Y L G L M L - - - - - H D A - - - L S N D -
SEQ D:	I T V - S Y L G L M L - - - - - H D A - - - L S N D -
SEQ E:	R N R N L I D G L M V - - - - - A - E E - C G D E

SEQ F: G K H L W N G G K K I V V T E G E I D M L T V M E L Q D C -

100

SEQ A:	- - - S V G G F A Y L A E L S K N T P S A A - N I S A Y A D
SEQ B:	- - - S V G G F A Y L A E L S K N T P S A A - N I S A Y A D
SEQ C:	K L R S V S H T V F L D D L S - - V C S A E E N L S N F I F
SEQ D:	K L R S V S H T V F L D D L S - - V C S A E E N L S N F I F
SEQ E:	- - - Y A T A V M M T A R A S - - C P S A A - N L K G Y A G

SEQ F: K Y P V V S L G H G - - - A S A A K K C A A - N Y E Y F D Q

125

SEQ A:	- - - - - I V R E R A V V R E M I S V
SEQ B:	- - - - - I V R E R A V V R D M I A V
SEQ C:	- - - - - R S F N E - Y N E N P L - -
SEQ D:	- - - - - R S F N E - Y N E N P L - -
SEQ E:	- - - - - M V A D S - Y Q R Q V L Q L
SEQ F:	F E Q I I L M F D M D E A G R K A V E E A A Q V L P A G K V

150

SEQ A:	A N E I A E A G F D P Q - - - - - G R T S E D L L D L A E S
SEQ B:	A H E I A D A G Y D P Q - - - - - G R N S D E L L D L A E S
SEQ C:	R R S P F L L L - E - R I K - - - - - G R - - - - - L D S A I A
SEQ D:	R R S P F L L L - E - R I K - - - - - G R - - - - - L D S A I A
SEQ E:	L D E M R E P I S N G T L D A S - G R - - - - - A M D E L V K
SEQ F:	R V - A V L P C K D A N E C H L N G H - - - - - D R E - -

175

Fragment II

SEQ A:	R V F K I A E S R - - A N K D E G P K N - I A D V L D A T V
SEQ B:	R V F Q I A E N R - - A N K D E G P K S - I D Q I L D A T V
SEQ C:	K T F S I R S A R G R S I Y D I F S Q S E I - G V L A R I K
SEQ D:	K T F S I R S A R G R S I Y D I F S Q S E I - G V L A R I K
SEQ E:	R L S S I R K P R N - E V K P V - R L G E I I N D Y T D T L
SEQ F:	- - - - I M E - - Q V W - N A - - - G P W I P D G V V S A L

200

SEQ A:	A R I E Q L F Q Q P H D G V T G V N - - - - - T G Y D D L
SEQ B:	A R I E Q L F Q Q P H D G V T G V D - - - - - T G Y Q D L
SEQ C:	K R R A A F S E N Q N S F F D G F P - - - - - T G Y K D I
SEQ D:	K R R A T F S E N Q N S F F D S F P - - - - - T G Y K D I
SEQ E:	D R R L - - - R N G E E S D T L K - - - - - T G I E E L
SEQ F:	S L E R E R I R E H L S S E E S V G L L F S G C T G I - - -

225

SEQ A:	N K K T A G L Q P S D L I I V A A R P S M G K T T F A M N L
SEQ B:	N K K T A G L Q R S D L I I V A A R P S M G K T T F A M N L
SEQ C:	D D K G V I L A K G N F V I I A A R P S I G K T A L A I D M
SEQ D:	D D K G V I L A K G N F V I I A A R P S I G K T A L A I D M
SEQ E:	D A I T G G M N A E D L V I I A A R P G M G K T E L A L K I
SEQ F:	N D K T L G A R G G E V I M T S G S G M G K S T F V R Q Q

250

SEQ A:	V	E	N	-	A	-	A	M	L	Q	D	K	-	-	P	-	-	-	-	-	-	V	L	I	F	S	L	E	M	
SEQ B:	C	E	N	-	A	-	A	M	L	Q	D	K	-	-	P	-	-	-	-	-	-	V	L	I	F	S	L	E	M	
SEQ C:	A	I	N	L	A	V	T	Q	-	Q	R	R	-	-	-	-	-	-	-	-	-	V	G	F	L	S	L	E	M	
SEQ D:	A	I	N	L	A	V	T	Q	-	Q	R	R	-	-	-	-	-	-	-	-	-	V	G	F	L	S	L	E	M	
SEQ E:	A	-	E	-	G	V	A	-	-	-	S	R	V	I	P	G	S	G	V	R	R	G	V	L	I	F	S	M	E	M

SEQ F: A L Q W G T A M G K K - - - - - V G L - A M E E

275

SEQ A:	P	S	E	Q	I	M	M	-	-	-	-	-	-	R	S	L	A	S	L	S	R	V	D	Q	T	-	-
SEQ B:	P	G	E	Q	I	M	M	-	-	-	-	-	-	R	M	L	A	S	L	S	R	V	D	Q	T	-	-
SEQ C:	S	A	G	Q	I	V	E	-	-	-	-	-	-	R	I	V	A	N	L	T	G	I	S	G	E	-	-
SEQ D:	S	A	G	Q	I	V	E	-	-	-	-	-	-	R	I	V	A	N	L	T	G	I	S	G	E	-	-
SEQ E:	S	A	I	Q	V	V	E	-	-	-	-	-	-	R	G	I	A	G	-	A	G	M	M	S	V	S	V

SEQ F: S V E E T A E D L I G L H N R V R L R Q S D S L K R E I I E

300

SEQ A:	K	-	I	R	T	G	Q	L	D	-	D	E	D	W	A	R	I	S	G	T	M	G	I	L	L	E	K	R	N	I
SEQ B:	R	-	I	R	T	G	Q	L	D	-	D	E	D	W	A	R	I	S	G	T	M	G	I	L	L	E	K	R	N	I
SEQ C:	K	-	L	Q	R	G	D	L	S	K	E	E	L	F	-	R	V	E	E	A	G	E	T	V	R	E	S	H	F	-
SEQ D:	K	-	L	Q	R	G	D	L	S	K	E	E	L	F	-	R	V	E	E	A	G	E	I	V	R	E	S	H	F	-
SEQ E:	R	N	P	S	R	M	D	-	-	-	D	E	G	W	A	R	V	A	S	G	M	K	L	L	A	E	L	D	V	-

SEQ F: K F D Q - W F - - - D E L F - - - - - G N - - - - D T F H L

325

SEQ A:	Y	I	D	D	S	S	G	L	T	P	T	E	V	R	S	R	A	R	R	I	A	R	-	E	H	G	G	I	G	L
SEQ B:	Y	I	D	D	S	S	G	L	T	P	T	E	V	R	S	R	A	R	R	I	F	R	-	E	H	G	G	L	S	L
SEQ C:	Y	I	C	S	D	S	Q	Y	K	L	N	L	I	-	A	N	Q	I	R	L	L	R	K	E	D	R	-	V	D	V
SEQ D:	Y	I	C	S	D	S	Q	Y	K	L	N	L	I	-	A	N	Q	I	R	L	L	R	K	E	D	R	-	V	D	V
SEQ E:	W	V	V	D	A	S	R	L	S	V	E	E	I	R	S	I	S	E	R	-	H	K	Q	E	H	P	N	L	S	L

SEQ F: Y - D S F A E A E T D R L L A K L A Y M R S G L G C D V - -

350

SEQ A:	I	M	I	D	Y	L	Q	L	M	-	R	V	P	A	L	S	D	N	R	T	L	E	I	A	E	I	S	R	S	L
SEQ B:	I	M	I	D	Y	L	Q	L	M	-	R	V	P	S	L	S	D	N	R	T	L	E	I	A	E	I	S	R	S	L
SEQ C:	I	F	I	D	Y	L	Q	L	I	-	N	-	S	S	V	G	E	N	R	Q	N	E	I	A	D	I	S	R	T	L
SEQ D:	I	F	I	D	Y	L	Q	L	I	-	N	-	S	S	V	G	E	N	R	Q	N	E	I	A	D	I	S	R	T	L
SEQ E:	I	M	A	D	Y	L	G	L	I	-	E	K	P	K	-	A	E	R	N	D	L	E	I	A	D	I	S	G	S	L

SEQ F: I I L D H I S I V V S A S G E S D E R - K M I D N L M T K L

Alignment of the amino acid sequence of *E. coli* dnaB protein and five other dnaB-like sequences.

Alignment of protein sequences was performed manually with reference to previous alignments (Wong *et al.*, 1988; Hatt *et al.*, 1988). Sources of the protein sequences were; SEQA = *E. coli* (Nakayama *et al.*, 1984b); SEQB = *S. typhimurium* (Wong *et al.*, 1988); SEQC = *C. trachomatis* (trachoma biovar; Sriprakash and Macavoy, 1987); SEQD = *C. trachomatis* (LGV biovar; Hatt *et al.*, 1988); SEQE = bacteriophage P22 (gene 12 protein; Backhaus and Petri, 1984); SEQF = bacteriophage T7 (gene 4 protein; Dunn and Studier, 1983).

Amino acid numbering is that of the *E. coli* sequence. Amino acid abbreviations are: A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine. Similarities between sequences are indicated by boxed amino acids: hatched = same amino acid; open = amino acids belonging to the same group (hydroxyl/small aliphatic: A, G, S, T; acid and acid amide: N, D, E, Q; basic: H, R, K; aliphatic: M, I, L, V; or aromatic: F, Y, W).

The enclosed sequences labelled Fragment II and Fragment III indicate those domain structures isolated following partial tryptic hydrolysis (Nakayama *et al.*, 1984a). The highlighted amino acid sequence I227→V243 of *E. coli* dnaB and related proteins indicates the proposed position of the ATP binding site in dnaB (Biswas *et al.*, 1986). The similarly highlighted amino acid sequence I361→L389 defines the leucine heptad repeat motif (Chapter 6). The sequence of bacteriophage T7 gene is separated from other sequences due to its low level of homology. This protein, however, does share extensive similarity with dnaB from *E. coli* and other sources in the region of the putative ATP binding site and sequences surrounding the leucine heptad repeat motif.

REFERENCES

- Alberts, B.M., Barry, J., Bedinger, P., Formosa, T., Jongeneel, C.V. and Kreuzer, K.N. (1982) Studies on DNA replication in the bacteriophage T4 *in vitro* system. *Cold Spring Harbor Symp. Quant. Biol.* **47**, 655-668
- Alfano, C. and McMacken, R. (1989) Ordered assembly of nucleoprotein structures at the Bacteriophage λ replication origin during the initiation of DNA replication. *J. Biol. Chem.* **264**, 10699-10708
- Allen, G.C.Jr. and Kornberg, A. (1991) The *priB* gene encoding the primosomal replication n protein of *Escherichia coli*. *J. Biol. Chem.* **266**, 11610-11613
- Andersson, S.G.E. and Kurland, C.G. (1990) Codon preferences in free living microorganisms. *Microbiol. Rev.* **54**, 198-210
- Anfinsen, C.B. (1973) Principles that govern the folding of protein chains. *Science* **181**, 223-230
- Arai, K.-I. and Kornberg, A. (1979) A general priming system employing only dnaB protein and primase for DNA replication. *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4308-4312
- Arai, K.-I. and Kornberg, A. (1981a) Unique primed start of phage ϕ X174 DNA replication and mobility of the primosome in a direction opposite chain synthesis. *Pro. Natl. Acad. Sci. U.S.A.* **78**, 69-73
- Arai, K.-I., Yasuda, S.-I. and Kornberg, A. (1981b) Mechanism of dnaB protein action. I. Crystallization and properties of dnaB protein, an essential replication protein in *Escherichia coli*. *J. Biol. Chem.* **256**, 5247-5252
- Arai, K.-I. and Kornberg, A. (1981c) Mechanism of dnaB protein action. II. ATP hydrolysis by dnaB protein dependent on single or double-stranded DNA. *J. Biol. Chem.* **256**, 5253-5259
- Arai, K.-I. and Kornberg, A. (1981d) Mechanism of dnaB protein action. III. Allosteric roles of ATP in the alteration of DNA structure by dnaB protein in priming replication. *J. Biol. Chem.* **256**, 5260-5266

- Arai, K.-I. and Kornberg, A. (1981e) Mechanism of dnaB protein action. IV. General priming of DNA replication by dnaB protein and primase compared with RNA polymerase. *J. Biol. Chem.* **256**, 5267-5272
- Arai, K.-I., Low, R., Kobori, J., Shlomai, J. and Kornberg, A. (1981f) Mechanism of dnaB protein action. V. Association of dnaB protein, protein n', and other prepriming proteins in the primosome of DNA replication. *J. Biol. Chem.* **256**, 5273-5280
- Arai, K.-I., McMacken, R., Yasuda, S.-I. and Kornberg, A. (1981g) Purification and properties of *Escherichia coli* protein i, a prepriming protein in ϕ X174 DNA replication. *J. Biol. Chem.* **256**, 5281-5286
- Atlung, T., Løbner-Olesen, A. and Hansen, F.G. (1987) Overproduction of dnaA protein stimulates initiation of chromosomal and minichromosome replication in *Escherichia coli*. *Mol. Gen. Genet.* **206**, 51-59
- Bachmann, B.J. (1990) Linkage map of *Escherichia coli* K-12. Edition 8. *Microbiol. Rev.* **54**, 130-197
- Backhaus, H. and Petri, J.B. (1984) Sequence analysis of a region from the early right operon in phage 22 including the replication genes 18 and 12. *Gene* **32**, 289-303
- Baker, T.A., Sekimizu, K., Funnel, B.E. and Kornberg, A. (1986) Extensive unwinding of the plasmid template during staged enzymatic initiation of DNA replication from the origin of the *Escherichia coli* chromosome. *Cell* **45**, 53-64
- Baker, T.A., Funnel, B.E. and Kornberg, A. (1987) Helicase action of dnaB protein during replication from the *Escherichia coli* chromosomal origin *in vitro*. *J. Biol. Chem.* **262**, 6877-6885
- Baker, T.A. and Kornberg, A. (1988) Transcriptional activation of initiation of replication from *E. coli* chromosomal origin: an RNA-DNA hybrid near *oriC*. *Cell* **55**, 113-123
- Barlow, D.J. and Thornton, J.M. (1988) Helix geometry in proteins. *J. Mol. Biol.* **201**, 601-619

- Bartholomew, B., Dahmus, M.E. and Meares, C.F. (1986) RNA contacts subunits Ilo and Ilc in HeLa RNA polymerase II transcription complexes. *J. Biol. Chem.* **261**, 14226-14231
- Berg, J.M. (1986) Potential metal-binding domains in nucleic acid binding proteins. *Science* **232**, 485-487
- Berg, J.M. (1990a) Zinc finger domains: hypotheses and current knowledge. *Annu. Rev. Biophys. Biophys. Chem.* **19**, 405-421
- Berg, J.M. (1990b) Zinc fingers and other metal-binding domains. Elements for interactions between macromolecules. *J. Biol. Chem.* **265**, 6513-6516
- Bernstein, J.A. and Richardson, C.C. (1988) A 7-kDa region of the bacteriophage T7 gene 4 protein is required for primase but not for helicase activity. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 396-400
- Bernstein, J.A. and Richardson, C.C. (1989) Characterization of the helicase and primase activities of the 63-kDa component of the bacteriophage T7 gene 4 protein. *J. Biol. Chem.* **264**, 13066-13073
- Biswas, E.E., Biswas, S.B. and Bishop, J.E. (1986) The dnaB protein of *Escherichia coli*: Mechanism of Nucleotide binding, hydrolysis and modulation by dnaC protein. *Biochemistry* **25**, 7368-7374
- Biswas, S.B. and Biswas, E.E. (1987) Regulation of dnaB function in DNA replication in *Escherichia coli* by dnaC and λ P gene products. *J. Biol. Chem.* **262**, 7831-7838
- Bittner, M. and Vapnek, D. (1981) Versatile cloning vectors derived from the runaway-replication plasmid pKN402. *Gene* **15**, 319-329
- Bordo, D. and Argus, P. (1991) Suggestions for "safe" residue substitutions in site directed mutagenesis. *J. Mol. Biol.* **217**, 721-729
- Bouché J.-P., Zechel, K. and Kornberg, A. (1975) dnaG gene product, a rifampicin-resistant RNA polymerase, initiates the conversion of a single-stranded coliphage DNA to its duplex replicative form. *J. Biol. Chem.* **250**, 5995-6001

- Bouché, J.-P., Rowen, L. and Kornberg, A. (1978) The RNA primer synthesized by primase to initiate phage G4 DNA replication. *J. Biol. Chem.* **253**, 765-769
- Boye, E. and Løbner-Olesen, A. (1990) The role of *dam* methyltransferase in the control of DNA replication in *E. coli*. *Cell* **62**, 981-989
- Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248-254
- Bramhill, D. and Kornberg, A. (1988a) Duplex opening by *dnaA* protein at novel sequences in initiation of replication at the origin of the *E. coli* chromosome. *Cell* **52**, 743-755
- Bramhill, D. and Kornberg, A. (1988b) A model for initiation at origins of DNA replication. *Cell* **54**, 915-918
- Braun, R.E., O'Day, K. and Wright, A. (1985) Autoregulation of the DNA replication gene *dnaA* in *E. coli* K-12. *Cell* **40**, 159-169
- Brinkmann, U., Mattes, R.E. and Buckel, P. (1989) High-level expression of recombinant genes in *Escherichia coli* is dependent on the availability of the *dnaY* gene product. *Gene* **85**, 109-114
- Burton, Z.F., Gross, C.A., Watanabe, K.K. and Burgess, R.R. (1983) The operon that encodes the sigma subunit of RNA polymerase also encodes ribosomal protein S21 and DNA primase in *E. coli* K12. *Cell* **32**, 335-349
- Campbell, J.L. and Kleckner, N. (1990) *E. coli oriC* and the *dnaA* gene promoter are sequestered from *dam* methyltransferase following the passage of the chromosomal replication fork. *Cell* **62**, 967-979
- Cha, T.-A. and Alberts, B.M. (1989) The bacteriophage T4 DNA replication fork. Only DNA helicase is required for leading strand DNA synthesis by the DNA polymerase holoenzyme. *J. Biol. Chem.* **264**, 12220-12225

- Chakerian, A.E., Tesmer, V.M., Manly, S.P., Brachett, J.K., Lynch, M.J., Hoh, J.T. and Matthews, K.S. (1991) Evidence for leucine zipper motif in lactose repressor protein. *J. Biol. Chem.* **266**, 1371-1374
- Challberg, M.D. and Kelly, T.S. (1989) Animal virus DNA replication. *Annu. Rev. Biochem.* **58**, 671-717
- Chambers, S.P., Prior, S.E., Barstow, D.A. and Minton, N.P. (1988) The pMTL *nic* cloning vectors. I. Improved pUC polylinker regions to facilitate the use of sonicated DNA for nucleotide sequencing. *Gene* **68**, 139-149
- Chang, A.C.Y. and Cohen, S.N. (1978) Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the PISA cryptic miniplasmid. *J. Bacteriol.* **134**, 1141-1156
- Chang, S.-F., Ng, D., Baird, L. and Georgopoulos, C. (1991) Analysis of an *Escherichia coli dnaB* temperature-sensitive insertion mutation and its cold sensitive extragenic suppressor. *J. Biol. Chem.* **266**, 3654-3660
- Chen, P.L. and Carl, P.L. (1975) Genetic map locations of the *Escherichia coli dnaG* gene. *J. Bacteriol.* **124**, 1613-1614
- Chou, P.Y. and Fasman, G.D. (1978) Empirical predictions of protein conformation. *Annu. Rev. Biochem.* **47**, 251-276
- Cohen, C. and Parry, D.A.D. (1986) α -helical coiled coils - a widespread motif in proteins. *TIBS* **11**, 245-248
- Coleman, J.E. (1983) The role of Zn(II) in RNA and DNA polymerases. In: *Zinc enzymes*. (Spiro, T.G., ed.), pp.219-252, John Wiley and Sons, New York
- Coleman, J. and Raetz, C.R.M. (1988) First committed step of LipidA biosynthesis in *Escherichia coli*: sequence of the *lpxA* gene. *J. Bacteriol.* **170**, 1268-1274
- Coleman, J.E. and Giedroc, D.P. (1989) Zinc-binding Proteins Involved in Nucleic Acid Replication. In: *metal ions in biological systems*. (Volume 25) (Sigel, H. and Sigel, A., eds.), pp.171-234, Marcel Dekker, Inc., New York
- Creighton, T.E. (1990) Protein Folding. *Biochem. J.* **270**, 1-16

- D'Arai, R. and Boulloc, P. (1990) Logic of the *Escherichia coli* cell cycle. *TIBS* **15**, 191-194
- Darst, S.A., Kubalek, E.W. and Kornberg, R.D. (1989) Three-dimensional structure of *Escherichia coli* RNA polymerase holoenzyme determined by electron crystallography. *Nature* **340**, 730-732
- Davis, R.W., Botstein, D. and Roth, J.R. (1980) *Advanced bacterial genetics*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
- deMassy, B., Bejar, S., Louarn, J., Louarn, J.-M. and Bouché, J.-P. (1987) Inhibition of replication forks exiting the terminus region of the *Escherichia coli* chromosome occurs at two loci separated by 5 min. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 1759-1763
- Dente, L., Cesareni, G. and Cortese, R. (1983) pEMBL: a new family of single-stranded plasmids. *Nucleic Acids Res.* **11**, 1645-1655
- DiGate, R.J. and Marians, K.J. (1988) Identification of a potent decatenating enzyme from *Escherichia coli*. *J. Biol. Chem.* **263**, 13366-13373
- Dixon, N.E., Gazzola, C., Blakely, R.L. and Zerner, B. (1976) Metal ions in enzymes using ammonia or amides. *Science* **191**, 1144-1150
- Dixon, N.E. and Sargeson, A.M. (1983) Roles for the metal ion in reactions of coordinated substrates and in some metalloenzymes. In: *Zinc enzymes*. (Spiro, T.G., ed.), pp.253-352, John Wiley and Sons, New York
- Dixon, N.E. and Kornberg, A. (1984) Protein HU in the enzymatic replication of the chromosomal origin of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **81**, 424-428
- Dodson, M., Roberts, J., McMacken, R. and Echols, H. (1985) Specialized nucleoprotein structures at the origin of replication of bacteriophage λ : complexes with λ O protein and with λ O, λ P and *Escherichia coli* dnaB proteins. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4678-4682

- Dodson, M., Echols, H., Wickner, S., Alfano, C., Mensa-Wilmot, K., Gomes, B., LeBowitz, J.H., Roberts, J.D. and McMacken, R. (1986) Specialized nucleoprotein structures at the origin of replication of bacteriophage λ : localized unwinding of duplex DNA by a six-protein reaction. *Pro. Natl. Acad. Sci. U.S.A.* **83**, 7638-7642
- Dodson, M., McMacken, R. and Echols, H. (1989) Specialized nucleoprotein structures at the origin of replication of bacteriophage λ . *J. Biol. Chem.* **264**, 10719-10725
- Dretzen, G., Bellard, M., Sassone-Corsi, P. and Chambon, P. (1981) A reliable method for the recovery of DNA fragments from agarose and acrylamide gels. *Anal. Biochem.* **112**, 295-298
- Dunn, J.J. and Studier, F.W. (1983) Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *J. Mol. Biol.* **166**, 477-535
- Ellis, R.J. and Hemmingsen, S.M. (1989) Molecular chaperones: proteins essential for the biogenesis of some macromolecular structures. *TIBS* **14**, 339-342
- Elvin, C.M., Dixon, N.E. and Rosenberg, H. (1986) Molecular cloning of the phosphate (inorganic) transport (*pit*) gene of *Escherichia coli* K12. Identification of the *pit*⁺ gene product and physical mapping of the *pit-gor* region of the chromosome. *Mol. Gen. Genet.* **204**, 477-484
- Elvin, C.M., Thompson, P.R., Argall, M.E., Hendry, P., Stamford, N.P.J., Lilley, P.E. and Dixon, N.E. (1990) Modified bacteriophage lambda promoter vectors for overproduction of proteins in *Escherichia coli*. *Gene* **87**, 123-126
- Engbaek, F., Gross, C.A. and Burgess, R.R. (1976) Quantitation of RNA polymerase subunits in *Escherichia coli* during exponential growth and after bacteriophage T4 infection. *Mol. Gen. Genet.* **143**, 291-295
- Erickson, B.D., Burton, Z.F., Watanabe, K.K. and Burgess, R.R. (1985) Nucleotide sequence of the *rpsU-dnaG-rpoD* operon from *Salmonella typhimurium* and a comparison of this sequence with the homologous operon of *Escherichia coli*. *Gene* **40**, 67-78

- Ernst, J.F. (1988) Codon usage and gene expression. *TIBtech.* **6**, 196-199
- Esaki, N. and Walsh, C.T. (1986) Biosynthetic alanine racemase of *Salmonella typhimurium*: Purification and characterization of the enzyme encoded by the *alr* gene. *Biochemistry* **25**, 3261-3267
- Fay, P.J., Johanson, K.O., McHenry, C.S. and Bambara, R.A. (1981) Size classes of products synthesized processively by DNA polymerase III and DNA polymerase III holoenzyme of *Escherichia coli*. *J. Biol. Chem.* **256**, 976-983
- Fay, P.J., Johanson, K.O., McHenry, C.S. and Bambara, R.A. (1982) Class sizes of products synthesized processively by two subassemblies of *Escherichia coli* DNA polymerase III holoenzyme. *J. Biol. Chem.* **257**, 5692-5699
- Ferrin, L.J., Mildvan, A.S. and Loeb, L.A. (1983) Metal content of DNA polymerase I purified from overproducing and wild type *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **112**, 723-728
- Firshein, W. (1989) Role of the DNA/membrane complex in prokaryotic DNA replication. *Annu. Rev. Microbiol.* **43**, 89-120
- François, V., Louarn, J. and Louarn, J.-M. (1989) The terminus of the *Escherichia coli* chromosome is flanked by several polar replication pause sites. *Mol. Microbiol.* **3**, 995-1002
- Fuller, R.S. and Kornberg, A. (1983) Purified dnaA protein in initiation of replication at the *E. coli* chromosomal origin of replication. *Proc. Natl. Acad. Sci. U.S.A.* **80**, 5817-5821
- Fuller, R.S., Funnell, B.E. and Kornberg, A. (1984) The dnaA protein complex with the *E. coli* chromosomal origin (*oriC*) and other DNA sites. *Cell* **38**, 889-900
- Funnell, B.E., Baker, T.A. and Kornberg, A. (1987) *In vitro* assembly of a prepriming complex at the origin of the *E. coli* chromosome. *J. Biol. Chem.* **262**, 10327-10334
- Galakatos, N.G., Daub, E., Botstein, D. and Walsh, C.T. (1986) Biosynthetic *alr* alanine racemase from *Salmonella typhimurium*: DNA and protein sequence determination. *Biochemistry* **25**, 3255-3260

- Garcia, G.M., Mar, P.K., Mullin, D.A., Walker, J.R. and Prather, N.E. (1986) The *E. coli* dnaY gene encodes an arginine transfer RNA. *Cell* **45**, 453-459
- Garnier, J., Osquithorpe, D.J. and Robson, B. (1978) Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.* **120**, 97-120
- Geider, K., Beck, E. and Schaller, H. (1978) An RNA transcribed from DNA at the origin of phage fd single strand to replicative form conversion. *Proc. Natl. Acad. Sci. U.S.A.* **75**, 645-649
- Gentz, R., Rauscher III, F.J., Abate, C. and Curran, T. (1989) Parallel association of Fos and Jun leucine zippers juxtaposes DNA binding domains. *Science* **243**, 1695-1699
- Georgopoulos, C.P. and Herskowitz, I. (1971) *Escherichia coli* mutants blocked in lambda DNA synthesis. In: *The bacteriophage lambda*. (Hershey, A.D., ed.), pp.553-564, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
- Gibson, F., Cox, G.B., Downie, J.A. and Radik, J. (1977) A mutation affecting a second component of the F_o portion of the magnesium ion-stimulated adenosine triphosphatase of *Escherichia coli* K12. The *uncC424* allele. *Biochem. J.* **164**, 193-198
- Giedroc, D.P. and Coleman, J.E. (1986) Structural and functional differences between the two intrinsic zinc ions of *Escherichia coli* RNA polymerase. *Biochemistry* **25**, 4969-4978
- Giedroc, D.P., Keating, K.M., Williams, K.R. and Coleman, J.E. (1987) The function of zinc in gene 32 protein from T4. *Biochemistry* **26**, 5251-5259
- Gill, S.C. and von Hippel, P.H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**, 319-326
- Gille, H. and Messer, W. (1991) Localized DNA melting and structural perturbations in the origin of replication, *oriC*, of *Escherichia coli* *in vitro* and *in vivo*. *EMBO J.* **10**, 1579-1584

- Giulian, G.G., Moss, R.L. and Greaser, M. (1983) Improved methodology for analysis and quantitation of proteins on one-dimensional silver-stained slab gels. *Anal. Biochem.* **129**, 277-287
- Godson, G.N. (1991) An over-expression plasmid for *Escherichia coli* primase. *Gene* **100**, 59-64
- Greenbaum, J.H. and Marians, K.J. (1984) The interactions of *Escherichia coli* replication factor Y with complementary strand origins of DNA replication. Contact points revealed by DNase footprinting and protection from methylation. *J. Biol. Chem.* **259**, 2594-2601
- Greenbaum, J.H. and Marians, K.J. (1985) Mutational analysis of primosome assembly sites. Evidence for alternative DNA structures. *J. Biol. Chem.* **260**, 12266-12267
- Grosjean, H. and Fiers, W. (1982) Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. *Gene* **18**, 199-209
- Günther, E., Mikalajczyk, M. and Schuster, H. (1981) Stabilization by ATP and ADP of *Escherichia coli* dnaB protein activity. *J. Biol. Chem.* **256**, 11970-11973
- Hatt, C., Ward, M.E. and Clarke, I.N. (1988) Analysis of the entire nucleotide sequence of the cryptic plasmid of *Chlamydia trachomatis* serovar LI. Evidence for involvement in DNA replication. *Nucleic Acids Res.* **16**, 4053-4067
- Hendry, P. and Sargeson, A.M. (1990) Metal ion promoted reactions of phosphate derivatives. In: *Progress in inorganic chemistry: bioinorganic chemistry*. (Lippard, S.J., ed.), **38**, pp.201-258, John Wiley and Sons, New York
- Henson, J.M., Chu, H., Irwin, C.A. and Walker, J.R. (1979) Isolation and characterization of dnaX and dnaY temperature-sensitive mutants of *Escherichia coli*. *Genetics* **92**, 1041-1059
- Hiasa, H., Sakai, H., Komano, T. and Godson, G.N. (1990) Structural features of the priming signal recognized by primase: mutational analysis of the phage G4 origin of complementary DNA strand synthesis. *Nucleic Acids Res.* **18**, 4825-4831

- Hill, T.M., Henson, J.M. and Kuempel, P.L. (1987) The terminus region of the *Escherichia coli* chromosome contains two separate loci that exhibit polar inhibition of replication. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 1754-1758
- Hill, T.M., Tecklenberg, M.L., Pelletier, A.J. and Kuempe, P.L. (1989) *tus*, the trans-acting gene required for termination of DNA replication in *Escherichia coli* encodes a DNA-binding protein. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 1593-1597
- Hill, T.M. and Marians, K.J. (1990) *Escherichia coli* Tus protein acts to arrest the progression of DNA replication forks *in vitro*. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 2481-2485
- Hines, J.C. and Ray, D.S. (1980) Construction and characterization of new coliphage M13 cloning vectors. *Gene* **11**, 207-218
- Hinkel, D.C. and Richardson, C.C. (1975) Bacteriophage T7 deoxyribonucleic acid replication *in vitro*. Purification and properties of the gene 4 protein of bacteriophage T7. *J. Biol. Chem.* **250**, 5523-5529
- Hinton, D.M., Richardson, R.W. and Nossal, N.G. (1987) Bacteriophage T4 DNA replication: Role of the T4 41 and 61 protein primase-helicase and the T4 uvsX recombination protein. In: *DNA replication and recombination*. (McMacken, R. and Kelly, T.J., eds.), pp.173-182, Alan R. Liss Inc., New York
- Hu, J.C., O'Shea, E.K., Kim, P.S. and Sauer, R.T. (1990) Sequence requirements for coiled-coils: analysis with λ repressor-GCN4 leucine zipper fusions. *Science* **250**, 1400-1403
- Hughes, P., Landoulsi, A. and Kohiyama, M. (1988) A novel role for cAMP in the control of the activity of the *E. coli* chromosome replication initiator protein, dnaA. *Cell* **55**, 343-350
- Hwang, D.S., Crooke, E. and Kornberg, A. (1990) Aggregated dnaA protein is dissociated and activated for DNA replication by phospholipase or dnaK protein. *J. Biol. Chem.* **265**, 19244-19248
- Hwang, D.S. and Kornberg, A. (1990) A novel protein binds a key origin sequence to block replication of an *E. coli* minichromosome. *Cell* **63**, 325-331

- Hwang, D.S. and Kaguni, J.M. (1991) dnaK protein stimulates a mutant form of dnaA protein in *Escherichia coli* DNA replication. *J. Biol. Chem.* **266**, 7537-7541
- Itaya, M. (1990) Isolation and characterization of a second RNase H (RNase HII) of *Escherichia coli* K-12 encoded by the *rnhB* gene. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 8587-8591
- Itoh, T. and Tomizawa, J. (1980) Formation of an RNA primer for initiation of replication of ColE1 DNA by ribonuclease H. *Proc. Natl. Acad. Sci. U.S.A.* **77**, 2450-2454
- Itoh, T. and Tomizawa, J. (1982) Purification of ribonuclease H as a factor required for initiation of *in vitro* ColE1 DNA replication. *Nucleic Acids Res.* **10**, 5949-5965
- Johnson, P.F. and McKnight, S.L. (1989) Eukaryotic transcriptional regulatory proteins. *Annu. Rev. Biochem.* **58**, 799-839
- Kane, J.F. and Hartley, D.L. (1988) Formation of recombinant protein inclusion bodies in *Escherichia coli*. *TIBtech.* **6**, 95-101
- Kaguni, J.M. and Kornberg, A. (1984) Replication initiated at the origin (*oriC*) of the *E. coli* chromosome reconstituted with purified proteins. *Cell* **38**, 183-190
- Katz, L., Kingsbury, D.T. and Helsinki, D.R. (1973) Stimulation by cyclic adenosine monophosphate of plasmid deoxyribonucleic acid replication and catabolite repression of the plasmid deoxyribonucleic acid-protein relaxation complex. *J. Bacteriol.* **114**, 577-591
- Khatri, G.S., MacAllister, T., Sista, P.R. and Bastia, D. (1989) The replication terminator protein of *E. coli* is a DNA sequence-specific contra-helicase. *Cell* **59**, 667-674
- King, G.C., Martin, C.T., Pham, T.T. and Coleman, J.E. (1986) Transcription by T7 RNA polymerase is not zinc-dependent and is abolished on amidomethylation of cysteine-347. *Biochemistry* **25**, 36-40
- Kingsbury, D.T. (1969) Estimate of the genome size of various microorganisms. *J. Bacteriol.* **98**, 1400-1401

- Kjeldgaard, N.O. and Gausing, K. (1974) Regulation of biosynthesis of ribosomes. In: *Ribosomes*. (Nomura, M., Tissieres, A. and Langyel, P., eds.), pp.369-392, Monograph Series, Cold Spring Harbor, New York
- Klein, A., Lanka, E. and Schuster, H. (1980) Isolation of a complex between the P protein at phage λ and the dnaB protein of *Escherichia coli*. *Eur. J. Biochem.* **105**, 1-6
- Kobayashi, T., Hidaka, M. and Horiuchi, T. (1989) Evidence of a *ter* specific binding protein essential for the termination reaction of DNA replication in *Escherichia coli*. *EMBO J.* **8**, 2435-2441
- Kobori, J.A. and Kornberg, A. (1982a) The *Escherichia coli* dnaC gene product. I. Overproduction of the dnaC proteins of *Escherichia coli* and *Salmonella typhimurium* by cloning into a high copy number plasmid. *J. Biol. Chem.* **257**, 13757-13762
- Kobori, J.A. and Kornberg, A. (1982b) The *Escherichia coli* dnaC gene product. II. Purification, physical properties, and role in replication. *J. Biol. Chem.* **257**, 13763-13769
- Kobori, J.A. and Kornberg, A. (1982c) The *Escherichia coli* dnaC gene product. III. Properties of the dnaB-dnaC protein complex. *J. Biol. Chem.* **257**, 13770-13775
- Konigsberg, W. and Godson, G.N. (1983) Evidence for use of rare codons in the *dnaG* gene and other regulatory genes of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **80**, 687-691
- Kornberg, A. (1982) *1982 supplement to DNA replication*. W.H. Freeman and Company, San Francisco
- Kornberg, A. and Baker, T. (1991) *DNA replication* (2nd edition) W.H. Freeman and Company, New York
- Kuempel, P.L., Pelletier, A.J. and Hill, T.M. (1989) Tus and the terminators: the arrest of replication in prokaryotes. *Cell* **59**, 581-583

- Kuempel, P.L., Pelletier, A.J. and Hill, T.M. (1990) *tus* and the terminators: inhibition of replication in *Escherichia coli*. In: *The Bacterial Chromosome*. (Drlica, K. and Riby, M., eds.), pp.297-306, American Society for Microbiology, Washington DC
- Kuramitsu, S., Inoue, K., Ogawa, T., Ogawa, H. and Kagamiyama, H. (1985) Aromatic amino acid aminotransferase of *Escherichia coli*: nucleotide sequence of the *TyrB* gene. *Biochem. Biophys. Res. Comm.* **133**, 134-139
- Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685
- Lambert, P.F., Waring, D.A., Wells, R.D. and Reznikoff, W.S. (1986) DNA requirements at the bacteriophage G4 origin of complementary-strand DNA synthesis. *J. Virol.* **58**, 450-458
- Landschulz, W.H., Johnson, P.F. and McKnight, S.L. (1988) The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* **240**, 1759-1764
- Landschulz, W.H., Johnson, P.F. and McKnight, S.L. (1989) The DNA binding domain of the rat liver nuclear protein C/EBP is bipartite. *Science* **243**, 1681-1688
- Lanka, E., Geschke, B. and Schuster, H. (1978) *Escherichia coli* *dnaB* mutant defective in DNA initiation: isolation and properties of the *dnaB* protein. *Proc. Natl. Acad. Sci. U.S.A.* **75**, 799-803
- Lanka, E. and Schuster, H. (1983) The *dnaC* protein of *Escherichia coli*. Purification, physical properties and interaction with *dnaB* protein. *Nucleic Acids Res.* **11**, 987-997
- Lark, K.G. (1972) Genetic control over the initiation of the synthesis of the short deoxynucleotide chains in *E. coli*. *Nature New Biol.* **240**, 237-240
- Lark, C.A., Riazi, J. and Lark, K.G. (1978) *dnaT*, dominant conditional-lethal mutation affecting DNA replication in *Escherichia coli*. *J. Biol. Chem.* **136**, 1008-1017

- Lasken, R.S. and Kornberg, A. (1988) The primosomal protein n' of *Escherichia coli* is a DNA helicase. *J. Biol. Chem.* **263**, 5512-5518
- LeBowitz, J.H., Zylicz, M., Georgopoulos, C. and McMacken, R. (1985) Initiation of DNA replication on single-stranded DNA templates catalyzed by purified replication proteins of bacteriophage λ and *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3988-3992
- LeBowitz, J.H. and McMacken, R. (1986) The *Escherichia coli* dnaB replication protein is a DNA helicase. *J. Biol. Chem.* **261**, 4738-4748
- Lee, E.H., Kornberg, A., Hidaka, M., Kobayashi, T. and Horiuchi, T. (1989) *Escherichia coli* replication termination protein impedes the action of helicases. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9104-9108
- Lee, E.H., Masai, H., Allen, G.C.Jr. and Kornberg, A. (1990) The *priA* gene encoding the primosomal replicative n' protein of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4620-4624
- Lee, E.H. and Kornberg, A. (1991) Replication deficiencies in *priA* mutants of *Escherichia coli* lacking the primosomal replication n' protein. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 3029-3032
- Lee, M.S. and Marians, K.J. (1987) *Escherichia coli* replication factor Y, a component of the primosome can act as a DNA helicase. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8345-8349
- Lee, M.S. and Marians, K.J. (1989) The *Escherichia coli* primosome can translocate actively in either direction along a DNA strand. *J. Biol. Chem.* **264**, 14531-14542
- Legerski, R.J. and Robberson, D.L. (1985) Analysis and optimization of recombinant DNA joining reactions. *J. Mol. Biol.* **181**, 297-312
- Liberek, K., Georgopoulos, C. and Zylicz, M. (1988) Role of the *Escherichia coli* dnaK and dnaJ heat shock proteins in the initiation of bacteriophage λ DNA replication. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6632-6636

- Loeber, G., Stenger, J.E., Ray, S., Parsons, R.E., Anderson, M.E. and Tegtmeyer, P. (1991) The zinc finger region of simian virus 40 large T antigen is needed for hexamer assembly and origin melting. *J. Virol.* **65**, 3167-3174
- Løbner-Olesen, A., Skarstad, K., Hansen, F.G., von Meyenburg, K. and Boye, E. (1989) The dnaA protein determines the initiation mass of *Escherichia coli* K-12. *Cell* **57**, 881-889
- Lothar, H., Kölling, R., Kucherer, C. and Schauzu, M. (1985) *dnaA* protein-regulated transcription: effects on the *in vitro* replication of *Escherichia coli* minichromosomes. *EMBO J.* **4**, 555-560
- Lupas, A., Van Dyke, M. and Stock, J. (1991) Predicting coiled coils from protein sequences. *Science* **252**, 1162-1164
- Lupski, J.R., Smiley, B.L. and Godson, G.N. (1983) Regulation of the *rpsU-dnaG-rpoD* macromolecular synthesis operon and the initiation of DNA replication in *Escherichia coli* K-12. *Mol. Gen. Genet.* **189**, 48-57
- Lupski, J.R. and Godson, G.N. (1984) The *rpsU-dnaG-rpoD* macromolecular synthesis operon of *E. coli*. *Cell* **39**, 251-252
- Luria, S.E. and Burrous, J.W. (1957) Hybridization between *Escherichia coli* and *Shigella*. *J. Bacteriol.* **74**, 461-476
- McHenry, C.S. (1988) DNA polymerase III holoenzyme of *Escherichia coli*. *Annu. Rev. Biochem.* **57**, 519-550
- McMacken, R., Bouché, J.-P., Rowen, S.L., Weiner, J.H., Ueda, K., Thelander, L., McHenry, C. and Kornberg, A. (1977) RNA priming of DNA replication. In: *Nucleic acid - protein recognition*. (Vogel, H.J., ed.), pp.15-29, Academic Press, London
- McMacken, R., Wold, M.S., LeBowitz, J.H., Roberts, J.D., Mallory, J.B., Wilkinson, J.A.K. and Loehrlein, C. (1983) Initiation of DNA replication *in vitro* promoted by the bacteriophage λ O and P replication proteins. In: *Mechanisms of DNA replication and recombination*. (Cozzarelli, N.R., ed.), pp.819-848, Alan R. Liss Inc., New York

- McMacken, R., Alfano, C., Gomas, B., LeBowitz, J.H., Mensa-Wilmot, K., Roberts, J.D. and Wold, M. (1986) Biochemical mechanisms in the initiation of bacteriophage λ DNA replication. In: *DNA replication and recombination*. (McMacken, R. and Kelly, T.J., eds.), pp.227-246, Alan R. Liss Inc., New York
- Maki, H., Maki, S. and Kornberg, A. (1988) DNA polymerase III holoenzyme of *Escherichia coli*. *J. Biol. Chem.* **263**, 6570-6578
- Mallory, J.B., Alfano, C. and McMacken, R. (1990) Host virus interactions in the initiation of bacteriophage lambda DNA replication. Recruitment of *Escherichia coli* dnaB helicase by lambda P replication protein. *J. Biol. Chem.* **265**, 13297-13307
- Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning. A laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
- Marston, F.A.O. and Hartley, D.L. (1990) Solubilization of protein aggregates. *Methods in Enzymol.* **182**, 264-276
- Masai, H., Bond, M.W. and Arai, K.-I. (1986) Cloning of the *Escherichia coli* gene for primosomal protein i: the relationship to *dnaT*, essential for chromosomal DNA replication. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 1256-1260
- Masai, H. and Arai, K.-I. (1988a) Initiation of lagging-strand synthesis for pBR322 plasmid DNA replication *in vitro* is dependent on primosomal protein i encoded by *dnaT*. *J. Biol. Chem.* **263**, 15016-15023
- Masai, H. and Arai, K.-I. (1988b) Operon structure of *dnaT* and *dnaC* genes essential for normal and stable DNA replication of *Escherichia coli* chromosome. *J. Biol. Chem.* **263**, 15083-15093
- Masai, H., Nomura, N., Kubota, Y. and Arai, K. (1990a) Roles of ϕ X174 type primosome- and G4 type primase-dependent primings in initiation of lagging and leading strand synthesis of DNA replication. *J. Biol. Chem.* **265**, 15124-15133
- Masai, H., Nomura, N. and Arai, K.-I. (1990b) The ABC-primosome. A novel priming system employing *dnaA*, *dnaB*, *dnaC* and primase on a hairpin containing a *dnaA* box sequence. *J. Biol. Chem.* **265**, 15134-15144

- Maurer, R., Osmond, B.C. and Botstein, D. (1984) Genetic analysis of DNA replication in bacteria: *dnaB* mutations that suppress *dnaC* mutations and *dnaQ* mutations that suppress *dnaE* mutations in *Salmonella typhimurium*. *Genetics* **108**, 25-38
- Maxon, M.E., Wigboldus, J., Brot, N. and Weissbach, H. (1990) Structure-function studies on *Escherichia coli* MetR protein, a putative prokaryotic leucine zipper protein. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 7076-7079
- Mead, D.A., Szczesna-Skorupa, E. and Kemper, B. (1986) Single-stranded DNA 'blue' T7 promoter plasmids: a versatile tandem promoter system for cloning and protein engineering. *Protein Engineering* **1**, 67-74
- Meyer, R.R., Glassberg, J. and Kornberg, A. (1979) An *Escherichia coli* mutant defective in single-strand binding protein is defective in DNA replication. *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1702-1705
- Meyer, R.R., Glassberg, J., Scott, J.V. and Kornberg, A. (1980) A temperature sensitive single-stranded DNA-binding protein from *Escherichia coli*. *J. Biol. Chem.* **255**, 2897-2901
- Mildvan, A.S. (1977) Magnetic resonance studies of the conformations of enzyme-bound substrates. *Acc. Chem. Res.* **10**, 246-252
- Miller, J., McLachlan, A.D. and Klug, A. (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J.* **4**, 1609-1614
- Minden, J.S. and Marians, K.J. (1985) Replication of pBR322 *in vitro* with purified proteins. Requirement for topoisomerase I in the maintenance of template specificity. *J. Biol. Chem.* **260**, 9316-9325
- Misra, R. and Reeves, P. (1985) Intermediates in the synthesis of TolC protein include an incomplete peptide stalled at a rare Arg codon. *Eur. J. Biochem.* **152**, 151-155
- Mizuuchi, K., Fisher, L.M., O'Dea, M. and Gellert, M. (1980) DNA gyrase action involves the introduction of transient double-strand breaks into DNA. *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1847-1851

- Mok, M. and Marians, K.J. (1987) The *Escherichia coli* preprimosome and dnaB helicase can form replication forks that move at the same rate. *J. Biol. Chem.* **262**, 16644-16654
- Monod, J., Cohen-Bazire, G. and Cohn, M. (1951) Sur la biosynthese de la β -galactosidase (lactase) chez *Escherichia coli*. La specificite de l'induction. *Biochem. Biophys. Acta.* **7**, 585-599
- Morrison, D.A. (1979) Transformation and preservation of competent bacterial cells by freezing. *Methods in Enzymol.* **68**, 326-331
- Mott, J.E., Grant, R.A., Hu, Y.-S. and Platt, T. (1985) Maximizing gene expression from plasmid vectors containing the λP_L promoter: strategies for overproducing transcription termination factor P. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 88-92
- Muramatsu, S. and Mizumo, T. (1990) Nucleotide sequence of the region encompassing the *int* gene of a cryptic prophage and the *dnaY* gene flanked by a curved DNA sequence of *Escherichia coli* K12. *Mol. Gen. Genet.* **220**, 325-328
- Murthy, M.R.N., Reid III, T.J., Sicignano, A., Tanaka, N. and Rossman, M.G. (1981) Structure of beef liver catalase. *J. Mol. Biol.* **152**, 465-499
- Nakai, H. and Richardson, C.C. (1988) Leading and lagging strand synthesis at the replication fork of bacteriophage T7. *J. Biol. Chem.* **263**, 9818-9830
- Nakayama, Y. and Yura, T. (1976) Induction of sigma factor synthesis in *Escherichia coli* by the λN gene product of bacteriophage lambda. *Proc. Natl. Acad. Sci. U.S.A.* **73**, 4405-4409
- Nakayama, N., Arai, N., Kaziro, Y. and Arai, K.-I. (1984a) Structural and functional studies of the dnaB protein using limited proteolysis. *J. Biol. Chem.* **259**, 88-96
- Nakayama, N., Arai, N., Bond, M.W., Kaziro, Y. and Arai, K.-I. (1984b) Nucleotide sequence of dnaB and the primary structure of the dnaB protein from *Escherichia coli*. *J. Biol. Chem.* **259**, 97-101

- Nash, H.A., Robertson, C.A., Flamm, E., Weisberg, R.A. and Miller, H.I. (1987) Overproduction of *Escherichia coli* integration host factor, a protein with nonidentical subunits. *J. Bacteriol.* **169**, 4124-4127
- Navaratnam, S., Myles, G.M., Strange, R.W. and Sancar, A. (1989) Evidence from extended X-ray absorption fine structure and site-specific mutagenesis for zinc fingers in *uvrA* protein of *Escherichia coli*. *J. Biol. Chem.* **264**, 16067-16071
- Norris, V. (1990) DNA replication in *Escherichia coli* is initiated by membrane detachment of *oriC*. A model. *J. Mol. Biol.* **215**, 67-71
- Nurse, P., DiGate, R.J., Zavitz, K.H. and Marians, K.J. (1990) Molecular cloning and DNA sequence analysis of *Escherichia coli priA*, the gene encoding the primosomal replication factor Y. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4615-4619
- Oas, T.G., McIntosh, L.P., O'Shea, E.K., Dahlquist, F.W. and Kim, P.S. (1990) Secondary structure of a leucine zipper determined by nuclear magnetic resonance spectroscopy. *Biochemistry* **29**, 2891-2894
- O'Donnell, M.E. and Kornberg, A. (1985) Dynamics of DNA polymerase III holoenzyme of *Escherichia coli* in replication of a multiprimed template. *J. Biol. Chem.* **260**, 12875-12883
- Ogasawara, N., Fujita, M.Q., Moriya, S., Fukuoka, T., Hirano, M. and Yoshikawa, H. (1990) Comparative anatomy of *oriC* of Eubacteria. In: *The Bacterial Chromosome*. (Drlica, K. and Riley, M., eds.), pp.282-696, American Society for Microbiology, Washington DC
- Ogawa, T., Baker, T.A., van der Ende, A. and Kornberg, A. (1985) Initiation of enzymatic replication at the origin of the *Escherichia coli* chromosome: contributions of RNA polymerase and primase. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3562-3566
- Oka, A., Sugimoto, K., Takanami, M. and Hirota, Y. (1980) Replication origin of the *E. coli* K-12 chromosome: the size and structure of the minimum DNA segment carrying information for autonomous replication. *Mol. Gen. Genet.* **178**, 9-20
- O'Shea, E.K., Rutkowski, R. and Kim, P.S. (1989a) Evidence that the leucine zipper is a coiled coil. *Science* **243**, 538-542

- O'Shea, E.K., Rutkowski, R., Stafford III, W.F. and Kim, P.S. (1989b) Preferential heterodimer formation by isolated leucine zippers from Fos and Jun. *Science* **245**, 646-648
- Plevani, P., Francesconi, S. and Lucchimi, G. (1987) The nucleotide sequence of the PRI1 gene related to DNA primase in *saccharomyces cerevisiae*. *Nucleic Acids Res.* **15**, 7975-7991
- Prussak, C.E., Almazan, M.T. and Tseng, B.Y. (1989) Mouse primase p49 subunit molecular cloning indicates conserved and divergent regions. *J. Biol. Chem.* **264**, 4957-4963
- Rao, R.N. and Rogers, S.G. (1978) A thermoinducible λ phage-ColE1 plasmid chimera for the overproduction of gene products from cloned DNA segments. *Gene* **3**, 247-263
- Rasmussen, R., Benvegna, D., O'Shea, E.K., Kim, P.S. and Alber, T. (1991) X-ray scattering indicates that the leucine zipper is a coiled coil. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 561-564
- Reha-Krantz, L.J. and Hurwitz, J. (1978) The *dnaB* gene product of *Escherichia coli*. I. Purification, homogeneity and physical properties. *J. Biol. Chem.* **253**, 4043-4050
- Richardson, C.C., Beauchamp, B.B., Huber, H.E., Ikeda, R.A., Meyers, J.A., Nakai, H., Rabkin, S.D., Tabor, S. and White, J. (1987) Bacteriophage T7 DNA replication. In: *DNA replication and recombination* (McMacken, R. and Kelly, T.J., eds.), pp151-171, Alan R. Liss Inc., New York
- Rosenberg, M. and Court, D. (1979) Regulatory sequences involved in the promotion and termination of RNA transcription. *Ann. Rev. Genet.* **13**, 319-353
- Rothman, J.E. (1989) Polypeptide chain binding proteins: catalysts for protein folding and related processes in cells. *Cell* **59**, 591-601
- Rowen, L. and Kornberg, A. (1978) A ribo-deoxyribonucleotide primer synthesized by primase. *J. Biol. Chem.* **253**, 770-774

- Rusche, J.R. and Howard-Flanders, P. (1985) Hexamine cobalt chloride promotes intermolecular ligation of blunt end DNA fragments by T4 DNA ligase. *Nucleic Acid Res.* **13**, 1997-2008
- Saito, H. and Uchida, H. (1977) Initiation of the DNA replication of bacteriophage lambda in *Escherichia coli* K12. *J. Mol. Biol.* **113**, 1-25
- Sakai, H. and Godson, G.N. (1985) Isolation and construction of mutants of the G4 minus strand origin: analysis of their *in vivo* activity. *Biochem. Biophys. Acta* **826**, 30-37
- Salas, M. (1988) Phages with protein attached to the DNA ends. In: *The Bacteriophages*. (Volume 1) (Calender, R., ed.), pp.169-191, Plenum Press, New York
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning. A laboratory manual*. Second edition. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463-5467
- Sasse-Dwight, S. and Gralla, J.D. (1990) Role of eukaryotic-type functional domains found in the prokaryotic enhancer receptor factor σ 54. *Cell* **62**, 945-954
- Saudek, V., Pastore, A., Catiglione Morelli, M.A., Frank, R., Gausephohl, H., Gibson, T., Weih, F. and Roesch, P. (1990) Solution structure of the DNA-binding domain of the yeast transcriptional activator protein GCN4. *Protein Engineering* **4**, 3-10
- Sayers, J.R., Schmidt, W. and Eckstein, F. (1988) 5'-3' exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis. *Nucleic Acids Res.* **16**, 791-802
- Schnier, J., Kitakawa, M. and Isono, K. (1986) The nucleotide sequence of an *Escherichia coli* chromosomal region containing the genes for ribosomal proteins S6, S18, L9 and an open reading frame. *Mol. Gen. Genet.* **204**, 126-132

- Schnoss, M., Zahn, K., Inman, R.B. and Blattner, R.R. (1988) Initiation protein induced helix destabilization at the λ origin: a prepriming step in DNA replication. *Cell* **52**, 385-395
- Schuermann, M., Hunter, J.B., Hennig, G. and Müller, R. (1991) Non-leucine residues in the leucine repeats of Fos and Jun contribute to the stability and determine the specificity of dimerization. *Nucleic Acids Res.* **19**, 739-746
- Sclafani, R.A. and Wechsler, J.A. (1981a) Suppression of *dnaC* alleles by the *dnaB* analog (*ban* protein) of bacteriophage P1. *J. Bacteriol.* **146**, 321-324
- Sclafani, R.A. and Wechsler, J.A. (1981b) Deoxyribonucleic acid initiation mutation *dnaB252* is suppressed by elevated *dnaC*⁺ gene dosage. *J. Bacteriol.* **146**, 412-418
- See, Y.P. and Jackowski, G. (1989) Estimating molecular weights of polypeptides by SDS gel electrophoresis. In: *Protein structure. a practical approach*. (Creighton, T.E., ed.), pp.1-21, IRL Press, Oxford
- Sekimizu, K., Bramhill, D. and Kornberg, A. (1987) ATP activates *dnaA* protein in initiating replication of plasmids bearing the origin of the *E. coli* chromosome. *Cell* **50**, 259-265
- Sekimizu, K. and Kornberg, A. (1988) Cardiolipin activation of *dnaA* protein, the initiation protein of replication in *E. coli*. *J. Biol. Chem.* **263**, 7131-7135
- Sekimizu, K., Yung, B.Y.-M. and Kornberg, A. (1988) The *dnaA* protein of *Escherichia coli*. *J. Biol. Chem.* **263**, 7136-7140
- Seufert, W. and Messer, W. (1987) Start sites for bidirectional *in vitro* DNA replication inside the replication origin *oriC* of *E. coli*. *EMBO J.* **6**, 2469-2472
- Sevastopoulos, C.G., Wehr, C.T. and Glaser, D.A. (1977) Large-scale automated isolation of *Escherichia coli* mutants with thermosensitive DNA-replication. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 3485-3489
- Shlomain, J. and Kornberg, A. (1980) A prepriming DNA replication enzyme of *Escherichia coli*. II. Actions of protein *n'*: a sequence specific, DNA dependent ATPase. *J. Biol. Chem.* **255**, 6794-6798

Shlomai, J., Polder, L., Arai, K.-I. and Kornberg, A. (1981) Replication of ϕ X174 DNA with purified enzymes. I. Conversion of viral DNA to a supercoiled biologically active duplex. *J. Biol. Chem.* **256**, 5233-5238

Silhavy, T.J., Berman, M.L. and Enquist, L.W. (1984) *Experiments with gene fusions*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Sista, P.R., Mukherjee, S., Patel, P., Khatri, G.S. and Bastia, D. (1989) A host-encoded DNA binding protein promotes termination of plasmid replication at a sequence-specific replication terminus. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 3026-3030

Skarstad, K., Baker, T.A. and Kornberg, A. (1990) Strand separation required for initiation of replication at the chromosomal origin of *E. coli* is facilitated by a distinct RNA-DNA hybrid. *EMBO J.* **9**, 2341-2348

Slaby, I., Lind, B. and Holmgren, A. (1984) T7 DNA polymerase is not a zinc-metalloenzyme and the polymerase and exonuclease activities are inhibited by zinc ions. *Biochem. Biophys. Res. Commun.* **122**, 1410-1417

Smiley, B.L., Lupski, J.R., Svec, P.S., McMacken, R. and Godson, G.N. (1982) Sequences of the *Escherichia coli* dnaG primase gene and regulation of its expression. *Proc. Natl. Acad. Sci. U.S.A.* **79**, 4550-4554

Solaiman, D. and Wu, F.Y.-H. (1985) Preparation and characterization of various *Escherichia coli* RNA polymerases containing one or two intrinsic metal ions. *Biochemistry* **24**, 5077-5083

Speckhard, D.C., Wu, F.Y.-H. and Wu, C.-W. (1977) Role of the intrinsic metal ion in RNA polymerase from *Escherichia coli*. *In vivo* substitution of tightly bound zinc with cobalt. *Biochemistry* **16**, 5228-5233

Sriprakash, K.S. and Macavoy, E.S. (1987) Characterization and sequence of a plasmid from the Trachoma biovar of *Chlamydia trachomatis*. *Plasmid* **18**, 205-214

Staudenbauer, W.L. (1977) Replication of the ampicillin resistance plasmid RSF 1030 in extracts of *Escherichia coli*: Separation of the replication cycle into early and late stages. *Mol. Gen. Genet.* **156**, 27-34

- Stayton, M.M. and Kornberg, A. (1983) Complexes of *Escherichia coli* primase with the replication origin of G4 phage DNA. *J. Biol. Chem.* **258**, 13205-13212
- Steck, T.R. and Drlica, K. (1984) Bacterial chromosome segregation: evidence for DNA gyrase involvement in decatenation. *Cell* **36**, 1081-1088
- Struhl, K. (1989) Helix-turn-helix, zinc-finger, and leucine zipper motifs for eukaryotic transcriptional regulatory proteins. *TIBS* **14**, 137-140
- Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods in Enzymol.* **185**, 60-88
- Stuitje, A.R., Weisbeek, P.J. and Meijer, M. (1984) Initiation signals for complementary strand DNA synthesis in the region of the replication origin of the *Escherichia coli* chromosome. *Nucleic Acids Res.* **12**, 3321-3332
- Stuitje, A.R., de Wind, N., van der Spek, J.C., Pors, T.H. and Meijer, M. (1986) Dissection of promoter sequences involved in transcriptional activation of the *Escherichia coli* replication origin. *Nucleic Acids Res.* **14**, 2333-2344
- Tabor, S. and Richardson, C.L. (1987) DNA sequence analysis with modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. U.S.A* **84**, 4767-4771
- Thomas, D.C., Kunkel, T.A., Casna, N.J., Ford, J.P. and Sancar, A. (1986) Activities and incision patterns of ABC exonuclease on modified DNA containing single-base mismatches and extrahelical bases. *J. Biol. Chem.* **261**, 14496-14505
- Tomasiewicz, H.G. and McHenry, C.S. (1987) Sequence analysis of the *Escherichia coli* *dnaE* gene. *J. Bacteriol.* **169**, 5735-5744
- Tse-Dinh, Y.-C. and Beran-Steed, R.K. (1988) *Escherichia coli* DNA topoisomerase I is a zinc metalloprotein with three repetitive zinc-binding domains. *J. Biol. Chem.* **263**, 15857-15859
- Tsurimoto, T., Hase, T., Matsubara, H. and Matsubara, K. (1982) Bacteriophage lambda initiators: preparation from a strain that overproduces the O and P proteins. *Mol. Gen. Genet.* **187**, 79-86

- Turner, R. and Tijian, R. (1989) Leucine repeats and an adjacent DNA binding domain mediate the formation of functional cFos-cJun heterodimers. *Science* **243**, 1689-1694
- Ueda, K., McMacken, R. and Kornberg, A. (1978) dnaB protein of *Escherichia coli*. *J. Biol. Chem.* **253**, 261-269
- van der Ende, A., Teertstra, R., van der Avoort, H.G.A.M. and Weisbeek, P.J. (1983) Initiation signals for complementary strand DNA synthesis on single-stranded plasmid DNA. *Nucleic Acids Res.* **11**, 4957-4975
- van der Ende, A., Baker, T.A., Ogawa, T. and Kornberg, A. (1985) Initiation of enzymatic replication at the origin of the *Escherichia coli* chromosome: primase as the sole priming enzyme. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3954-3958
- van Houten, B. (1990) Nucleotide excision repair in *Escherichia coli*. *Microbiol. Rev.* **54**, 18-51
- Vieira, J. and Messing, J. (1982) The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* **19**, 259-268
- Vieira, J. and Messing, J. (1987) Production of single-stranded plasmid DNA. *Methods in Enzymol.* **153**, 3-11
- von Meyenburg, K. and Hansen, F.G. (1987) Regulation of Chromosome Replication. In: *Escherichia coli and Salmonella typhimurium: cellular and molecular biology*. (Volume 2) (Neidhardt, F.D., Ingraham, J.L., Low, K.B., Magasanik, B., Schaechter, M. and Umberger, H.E., eds.), pp.1555-1577, American Society for Microbiology, Washington DC
- Wada, K., Aota, S., Tsuchiya, R., Ishibashi, F., Gojobori, T. and Ikemura, T. (1990) Codon usage tabulated from the Genbank genetic sequence data. *Nucleic Acids Res.* **18**, 2367-2411
- Wahle, E., Lasken, R.S. and Kornberg, A. (1989a) The dnaB-dnaC replication protein complex of *Escherichia coli*. I. Formation and properties. *J. Biol. Chem.* **264**, 2463-2468

- Wahle, E., Lasken, R.S. and Kornberg, A. (1989b) The dnaB-dnaC replication protein complex of *Escherichia coli*. II. Role of the complex in mobilizing dnaB functions. *J. Biol. Chem.* **264**, 2469-2475
- Wang, L.-F., Price, C.W. and Doi, R.H. (1985) *Bacillus subtilis* dnaE Encodes a Protein Homologous to DNA Primase of *Escherichia coli*. *J. Biol. Chem.* **260**, 3368-3372
- Wang, J.C. (1991) DNA topoisomerases: why so many? *J. Biol. Chem.* **266**, 6659-6662
- Ward, D.F. and Gottesmann, M.E. (1982) Suppression of transcription termination by phage λ . *Science* **216**, 946-951
- Wechsler, J.A. and Gross, J.D. (1971) *E. coli* mutants temperature-sensitive for DNA synthesis. *Mol. Gen. Genet.* **113**, 273-284
- Wechsler, J.A. (1975) Genetic and phenotypic characterization of dnaC mutations. *J. Bacteriol.* **121**, 594-599
- Wickner, S. and Hurwitz, J. (1975) Interaction of *Escherichia coli* dnaB and dnaC(D) gene products *in vitro*. *Proc. Natl. Acad. Sci. U.S.A.* **72**, 921-925
- Wickner, S.H. (1979) DNA replication proteins of *Escherichia coli* and phage λ . *Cold Spring Harbor Symp. Quant. Biol.* **43**, 303-310
- Wijsman, H.J.W. (1972) The characterization of an alanine racemase mutant of *Escherichia coli*. *Genet. Res.* **20**, 269-277
- Wold, M.S. and McMacken, R. (1982) Regulation of expression of the *Escherichia coli* dnaG gene and amplification of the dnaG primase. *Proc. Natl. Acad. Sci. U.S.A.* **79**, 4907-4911
- Wong, A., Kean, L. and Maurer, R. (1988) Sequence of the dnaB gene of *Salmonella typhimurium*. *J. Bacteriol.* **170**, 2668-2675

- Wu, F.Y.-H. and Wu, C.-W. (1983) The role of zinc in DNA and RNA polymerases. In: *Metal ions in biological systems: zinc and its role in biology and nutrition*. (Volume 15) (Sigel, H., ed.), pp.157-192, Marcel Dekker Inc., New York
- Wu, F.Y.-H. (1986) Biochemical and Biophysical studies on the intrinsic zinc ions in *E. coli* RNA polymerase. In: *Zinc enzymes*. (Volume 1) (Bertini, I., Luchinat, C., Maret, W. and Zeppezaur, M., eds.), pp.535-551, Birkhauser, Basel
- Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* 33, 103-119
- Yung, B.Y.-M. and Kornberg, A. (1988) Membrane attachment activates dnaA protein, the initiation protein of chromosomal replication in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 85, 7202-7205
- Yung, B.Y.-M. and Kornberg, A. (1989) The dnaA initiator protein binds separate domains in the replication origin of *Escherichia coli*. *J. Biol. Chem.* 264, 6146-6150
- Zavitz, K.H., DiGate, R.J. and Marians, K.J. (1991) The priB and priC replication proteins of *Escherichia coli*. Genes, DNA sequence, overexpression and purification. *J. Biol. Chem.* 266, 13988-13995
- Zipursky, S.L. and Marians, K.J. (1980) Identification of two *Escherichia coli* factor Y effector sites near the origins of replication of the plasmids ColE1 and pBR322. *Proc. Natl. Acad. Sci. U.S.A.* 77, 6521-6525
- Zipursky, S.L. and Marians, K.J. (1981) *Escherichia coli* factor Y sites of plasmid pBR322 can function as origins of DNA replication. *Proc. Natl. Acad. Sci. U.S.A.* 78, 6111-6115
- Zylicz, M., Gorska, I., Taylor, K. and Georgopoulos, C. (1984) Bacteriophage λ replication proteins: formation of a mixed oligomer and binding to the origin of λ DNA. *Mol. Gen. Genet.* 196, 401-406

Zylicz, M., Ang, D., Liberak, K., Yamamoto, T. and Georgopoulos, C. (1988) Initiation and lambda DNA replication reconstituted with purified lambda and *Escherichia coli* replication proteins. *Biochem. Biophys. Acta.* **951**, 344-350

Zylicz, M., Ang, D., Liberek, K. and Georgopoulos, C. (1989) Initiation of λ DNA replication with purified host- and bacteriophage-encoded proteins: the role of the dnaK, dnaJ and grpE heat shock proteins. *EMBO J.* **8**, 1601-1608

Zyskind, J.W. and Smith, D.W. (1977) Novel *Escherichia coli* *dnaB* mutant: direct involvement of the *dnaB252* gene product in the synthesis of an origin-ribonucleic acid species during initiation of a round of deoxyribonucleic acid replication. *J. Bacteriol.* **129**, 1476-1486

Zyskind, J.W. and Smith, D.W. (1986) The bacterial origin of replication, *oriC*. *Cell* **46**, 489-490

Zyskind, J.W. (1990) Priming and growth rate regulation: questions concerning initiation of DNA replication in *Escherichia coli*. In: *The bacterial chromosome*. (Drlica, K. and Riley, M., eds.), pp.268-278, American Society for Microbiology, Washington DC